Lecture Notes in Networks and Systems 322

Mostafa Al-Emran
Mohammed A. Al-Sharafi
Mohammed N. Al-Kabi
Khaled Shaalan *Editors*

Proceedings of International Conference on Emerging Technologies and Intelligent Systems

ICETIS 2021 Volume 2

Springer

Lecture Notes in Networks and Systems

Volume 322

Series Editor

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Advisory Editors

Fernando Gomide, Department of Computer Engineering and Automation—DCA, School of Electrical and Computer Engineering—FEEC, University of Campinas—UNICAMP, São Paulo, Brazil

Okyay Kaynak, Department of Electrical and Electronic Engineering, Bogazici University, Istanbul, Turkey

Derong Liu, Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, USA,
Institute of Automation, Chinese Academy of Sciences, Beijing, China

Witold Pedrycz, Department of Electrical and Computer Engineering, University of Alberta, Alberta, Canada,
Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

Marios M. Polycarpou, Department of Electrical and Computer Engineering, KIOS Research Center for Intelligent Systems and Networks, University of Cyprus, Nicosia, Cyprus

Imre J. Rudas, Óbuda University, Budapest, Hungary

Jun Wang, Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

Indexed by SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

More information about this series at https://link.springer.com/bookseries/15179

Mostafa Al-Emran · Mohammed A. Al-Sharafi ·
Mohammed N. Al-Kabi · Khaled Shaalan
Editors

Proceedings of International Conference on Emerging Technologies and Intelligent Systems

ICETIS 2021 Volume 2

Springer

Editors
Mostafa Al-Emran
The British University in Dubai
Dubai, United Arab Emirates

Mohammed A. Al-Sharafi
Universiti Teknologi Malaysia
Johor, Malaysia

Mohammed N. Al-Kabi
Al Buraimi University College
Al Buraimi, Oman

Khaled Shaalan
The British University in Dubai
Dubai, United Arab Emirates

ISSN 2367-3370 ISSN 2367-3389 (electronic)
Lecture Notes in Networks and Systems
ISBN 978-3-030-85989-3 ISBN 978-3-030-85990-9 (eBook)
https://doi.org/10.1007/978-3-030-85990-9

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Scientific Committee

Conference General Chair

Prof. Khaled Shaalan, The British University in Dubai, United Arab Emirates.

Conference Organizing Chair

Dr. Mostafa Al-Emran, The British University in Dubai, United Arab Emirates.

Program Committee Chairs

Dr. Mohammed A. Al-Sharafi, Universiti Teknologi Malaysia, Malaysia.
Dr. Mohammed N. Al-Kabi, Al Buraimi University College, Oman.

Publication Committee Chairs

Dr. Mohammed A. Al-Sharafi, Universiti Teknologi Malaysia, Malaysia.
Dr. Mostafa Al-Emran, The British University in Dubai, United Arab Emirates.

Conference Tracks Chairs

Prof. Vitaliy Mezhuyev, FH JOANNEUM University of Applied Sciences, Austria.
Dr. Garry Tan Wei Han, UCSI University, Malaysia.

Dr. Ibrahim Arpaci, Gaziosmanpaşa University, Turkey.
Dr. AbdulRahman Al-Sewari, Universiti Malaysia Pahang, Malaysia.
Dr. Baraq Ghaleb, Edinburgh Napier University, UK.
Dr. Abdullah Nasser, Universiti Malaysia Pahang, Malaysia.

Members of Scientific Committee

Abdallah Namoun, Islamic University of Medina, Saudi Arabia.
Abdualmajed Al-Khulaidi, Sana'a University, Yemen.
Abdulmajid Mohammed Abdulwahab Aldaba, International Islamic University Malaysia, Malaysia.
Abeer Dawood Salman Al-Nuaimi, Al Maarif University College, Iraq.
Ahmad Mohd Aziz Hussein, Umm Al-Qura Universit, Saudi Arabia.
Ahmed Jamal Ahmed, Al Maarif University College, Iraq.
Ahmed M. Mutahar, Management and Science University, Malaysia.
Aisyah Ibrahim, Universiti Malaysia Pahang, Malaysia.
Akhyari Nasir, University College TATI, Kemaman, Terengganu, Malaysia.
Alaa A. D. Taha, University of Mosul, Iraq.
Ali Ameen, Lincoln University, Malaysia.
Ali Nasser Ali Al-Tahitah, Universiti Sains Islam Malaysia, Malaysia.
Ali Qasem Saleh Al-Shetwi, Universiti Tenaga Nasional, Malaysia.
Ameen A. Ba Homaid, Universiti Malaysia Pahang, Malaysia.
Amir A. Abdulmuhsin, University of Mosul, Iraq.
Amr Abdullatif Yassin, Ibb University, Yemen.
Arwa Y. Aleryani, Independent Researcher, Canada
Azliza Yacob, University College TATI, Kemaman, Terengganu, Malaysia.
Bander Ali Saleh Al-Rimy, Universiti Teknologi Malaysia, Malaysia.
Basheer Mohammed Al-Haimi, Hebei University, Boading, China.
Bokolo Anthony Jnr., Norwegian University of Science and Technology, Norway.
Dalal Abdulmohsin Hammood, Middle Technical University, Iraq.
Eissa M. Alshari, Universiti Malaysia Kelantan, Malaysia.
Fadi A. T. Herzallah, Palestine Technical University, Kadoorie, Palestine.
Fathey Mohammed, Universiti Utara Malaysia, Malaysia.
Fuad Ali Mohammed Al-Yarimi, King Khalid University, Saudi Arabia.
Ghaleb H. Al-Gaphari, Sana'a University, Yemen.
Ghassan Saleh Hussein ALDharhani, Geomatika University College, Malaysia
Gonçalo Marques, Universidade da Beira Interior, Portugal.
Heider A. M. Wahsheh, King Faisal University, Saudi Arabia.
Hussam S. Alhadawi, Dijlah University College, Iraq.
Hussein Mohammed Esmail Abu Al-Rejal, University Utara Malaysia, Malaysia.
Ismail Ahmed Al-Qasem Al-Hadi, Universiti Malaysia Terengganu, Malaysia.
Joseph Ng, UCSI University, Malaysia.
Joshua A. Abolarinwa, Namibia University of Science and Technology, Namibia.

Kamal Ali Alezabi, UCSI University, Malaysia.
Kamal Mohammed Alhendawi, Al-Quds Open University. Faculty of Management, Palestine.
Kauthar Mohd Daud, Universiti Malaysia Kelantan, Malaysia.
Majid Abdulateef Abdullah, University of Hafr Al Batin, Saudi Arabia.
Majid Salal Naghmash, Dijlah University College, Iraq.
Mohamed Elwakil, University of Cincinnati, USA.
Mohammed Ahmed Talab, Al Maarif University College, Iraq.
Muaadh Mukred, Universiti Kebangsaan Malaysia, Malaysia.
Muhammad Akmal Remli, Universiti Malaysia Kelantan, Malaysia.
Mustafa Musa Jaber, Dijlah University College, Iraq.
Nejood Hashim Al-Walidi, Sanaa University, Yemen.
Noor Akma Abu Bakar, Universiti Malaysia Pahang, Malaysia.
Noor Al-Qaysi, Universiti Pendidikan Sultan Idris, Malaysia.
Noor Suhana Sulaiman, University College TATI, Kemaman, Terengganu, Malaysia.
Nooraida Samsudin, University College TATI, Kemaman, Terengganu, Malaysia.
Nur Ilyana Binti Ismarau Tajuddin, UCSI University, Malaysia.
Nur Sukinah Aziz, University College TATI, Kemaman, Terengganu, Malaysia.
Osama Al Sayaydeh, Universiti Malaysia Pahang, Malaysia.
Osama Mohammad Aljarrah, University of Massachusetts Dartmouth, USA.
Osamah A. M. Ghaleb, Almustaqbal University, Saudi Arabia.
Qasem Al-Tashi, Universiti Teknologi PETRONAS, Malaysia.
Qasim Al-Ajmi, A' Sharqiyah University, Oman.
Rayan Alanazi, Jouf University, Saudi Arabia.
Redhwan Al-Amri, Taylor's University, Malaysia.
Safwan Mahmood Yahya Ghaleb, University Malaysia Terengganu, Malaysia.
Salim M. Zaki, Dijlah University College, Iraq.
Samer Ali Alshami, Universiti Teknikal Malaysia Melaka, Malaysia.
Shahad Al-Yousif, Dijlah University College, Iraq.
Sinan Salih, Dijlah University College, Iraq.
Suhaila Zainudin, Universiti Kebangsaan Malaysia, Malaysia.
Sura Khalil Abd, Dijlah University College, Iraq.
Syahrul Fahmy Abdul Wahab, University College TATI, Kemaman, Terengganu, Malaysia.
Taha Hussein Rassem, Universiti Malaysia Pahang, Malaysia.
Tuty Asmawaty Abdul Kadir, Universiti Malaysia Pahang, Malaysia.
Vitaliy Mezhuyev, FH JOANNEUM University of Applied Sciences, Austria.
Waheed Ali Hussein Mohammed Ghanem, Universiti Malaysia Terengganu, Malaysia.
Yahya M. Al-Dheleai, Universiti Sains Malaysia, Malaysia.

Publicity Committee

Redhwan Al-Amri, Taylor's University, Kuala Lumpur, Malaysia.
Noor Akma Abu Bakar, Universiti Malaysia Pahang, Kuantan, Malaysia.

Preface

During the last decade, nobody denies the role of emerging technologies and intelligent systems in building digital societies and reshaping the knowledge-based economy. The number of experimental and theoretical findings is increasing rapidly due to the large number of successful emerging technologies. The role of those technologies has been witnessed across various domains, including healthcare, education, tourism, and industry, among others. The aim of the International Conference on Emerging Technologies and Intelligent Systems (ICETIS 2021) is to provide a forum for academics, scholars, and practitioners from academia and industry to share and exchange the recent developments in the domain of intelligent systems and emerging technologies. The ICETIS 2021 also aims to shed light on the emerging research trends in intelligent systems and their applications. Further, it attempts to concentrate on the recent design, developments, and modifications of intelligent systems and their applications. It also aims to provide a holistic view of the main theoretical models used to understand the adoption, acceptance, and continued use of intelligent systems and their applications.

The ICETIS 2021 was able to attract 249 submissions from 30 different countries across the globe. From the 249 submissions, we accepted 139 submissions, which represents an acceptance rate of 55.8%. Out of the 139 accepted submissions, 82 were selected to be published in this volume. The accepted papers were categorized under four different themes, including *Artificial Intelligence and Soft Computing*, *Information Security and Networking*, *Medical Informatics*, and *Advances in Information Systems*. Each submission is reviewed by at least two reviewers, who are considered experts in the related submitted paper. The evaluation criteria include several issues, such as correctness, originality, technical strength, significance, quality of presentation, interest, and relevance to the conference scope. The conference proceedings are published in *Lecture Notes in Networks and Systems Series* by Springer, which has a high SJR impact.

We acknowledge all those who contributed to the staging of the ICETIS 2021. We would also like to express our gratitude to the reviewers for their valuable feedback and suggestions. Without them, it was not possible to maintain the high quality and the success of the ICETIS 2021. As gratitude for their efforts, ICETIS

2021 is partnered with Publons to give an official recognition for the reviewers' contribution to peer review. This partnership means that reviewers can opt-in to have their reviews added to their Publons profile.

Dubai, United Arab Emirates — Mostafa Al-Emran
Johor, Malaysia — Mohammed A. Al-Sharafi
Al Buraimi, Oman — Mohammed N. Al-Kabi
Dubai, United Arab Emirates — Khaled Shaalan

Contents

Artificial Intelligence and Soft Computing

ANN-Based Reference Evapotranspiration Estimation: Effects of Data Normalization and Parameters Selection 3
Min Yan Chia, Yuk Feng Huang, and Chai Hoon Koo

Trend Analysis of the Effects of Climate Change on the Pan Evaporation Rate in Sabah, Malaysia 13
Min Yan Chia, Yuk Feng Huang, Chai Hoon Koo, and Yeu Zhi Tan

Cow Image Segmentation Using Mask R-CNN Integrated with Grabcut 23
Rotimi-Williams Bello, Ahmad Sufril Azlan Mohamed, and Abdullah Zawawi Talib

Forecasting of Reservoir Inflow Using Machine Learning—Case Study: Klang Gate Dam Reservoir 33
Karim Sherif Mostafa Hassan, Yuk Feng Huang, Chai Hoon Koo, Tan Kok Weng, Ali Najah Ahmed, and Ahmed Hussein Kamel Ahmed Elshafie

An Unsupervised Semantic Model for Arabic/French Terminology Extraction 49
Setha Imene and Aliane Hassina

Research on Improving E-commerce Logistics Service Customer Satisfaction Through Application of AGV in Intelligent Warehouse 61
Ying Xi, Noorsiah Binti Ahmad, and Abdullah Al Mamun

Analyzing Tweet Data to Identify the Impact of Demonetarization 75
Kannan Suseelan Unnithan, Said A. Salloum, and Khaled Shaalan

Qualitative Study in Natural Language Processing: Text Classification 83
Ahlam Wahdan, Said A. Salloum, and Khaled Shaalan

Heterogeneous Neural Collaborative Filtering for a Business Recommender System 93
Vu Hong Quan and Phan Duy Hung

An Online Clustering Approach for Evolving Data-Stream Based on Data Point Density 105
Md Kamrul Islam and Bishnu Sarker

Improving Sentiment Classification for Large-Scale Social Reviews Using Stack Generalization 117
Lamees Abu Romman, Shaheen Khatoon Syed, Majed Alshmari, and Md Maruf Hasan

A Comparative Study on Autism Among Children Using Machine Learning Classification 131
Ainie Hayati Binti Noruzman, Ngahzaifa Ab Ghani, and Nor Saradatul Akmar Zulkifli

The Role of Gear Grind Games in Creating a Motivating Learning Environment 141
W. L. WilliamCheng, P. S. JosephNg, H. C. Eaw, and K. Y. Phan

Sensitivity Analysis of Intelligent Active Force Control Applied to a Quadrotor System 153
Sherif I. Abdelmaksoud, Musa Mailah, and Ayman M. Abdallah

Scalable Smartification of Commercial Buildings HVAC Systems Using the Internet of Things and Machine Learning 165
Mohammad Najah Mahdi, Taofiq Adeola Bakare, Abdul Rahim Ahmad, Adamu Muhammad Buhari, and Khalid Sheikhidris Mohamed

Features Extraction for Robust Face Recognition Using GLCM and CS-LBP 175
Abeer D. Salman, Mohammed Ahmed Talab, and Ruqayah R. Al-Dahhan

Toward a Full Exploitation of IoT in Smart Cities: A Review of IoT Anomaly Detection Techniques 193
Redhwan Al-Amri, Raja Kumar Murugesan, Eissa M. Alshari, and Hussam S. Alhadawi

Gaze Prediction Based on Convolutional Neural Network 215
Abdulkader Helwan, Mohammad Khaleel Sallam Ma'aitah, Selin Uzelaltinbulat, Mohamad Ziad Altobel, and Manal Darwish

Stacked Autoencoders Deep Learning Approach for Left Ventricular Localization in Magnetic Resonance Slices 225
Abdulkader Helwan, Mohammad Khaleel Sallam Ma'aitah, Selin Uzelaltinbulat, Bengi Sonyel, Mohamad Ziad Ziad Altobel, and Manal Darwish

Machine Learning Based Prediction of Edematous Malnutrition in Afghan Children 235
Ziaullah Momand, Mohammad Shuaib Zarinkhail, and Mohammad Farhad Aryan

Discrete Particle Swarm Optimization for User Grouping in 5G NOMA Networks 247
Hadhrami Ab. Ghani, Farah Najwa Roslim, Eissa Mohammed Mohsen Al-Shari, Muhammad Akmal Remli, Nurul Izrin Md Saleh, Nooraini Yusoff, Hasyiya Karimah Adli, and Khairul Nizar Syazwan Wan Salihin Wong

Ant-Colony Based User Grouping Modelling for Fast Cellular Networks 259
Ahmed Abdullah Ahmed Alhanik, Nur Asyiqin Hamzah, Hadhrami Ab. Ghani, Mohammed Ahmed Salem, Azlan Abdul Aziz, and Eissa Mohammed Mohsen Al-Shari

A Systematic Approach to Transform Machine Learning Students' Performance Prediction Model into Preventive Procedures 269
Ijaz Khan, Abdul Rahim Ahmad, Nafaa Jabeur, and Mohammed Najah Mahdi

A Review on Intrusion Detection System Using a Machine Learning Algorithms 281
Huda Sirag and Sally Dafaallah Awadelkariem

A Systematic Review for Sentiment Analysis of Arabic Dialect Texts Researches 291
Arwa A. Al Shamsi and Sherief Abdallah

Crowdsourced NLP Retraining Engine in Chatbots 311
Mohammad Arif Nezami and Rakhshanda Rukham

Early Detecting Students at Risk Using Machine Learning Predictive Models 321
Ahlam Wahdan, Sendeyah Hantoobi, Mostafa Al-Emran, and Khaled Shaalan

Chaotic Particle Swarm Optimization Based on Meeting Room Approach for Designing Bijective S-Boxes 331
Hussam S. Alhadawi, Sinan Q. Salih, and Yasir Dawood Salman

Information Security and Networking

Performance of Effect for XPM and FWM in Fiber Optics 345
Mohammed Salah Abood, Mustafa Maad Hamdi, Bakr Ahmed Taha, Ahmed Shamil Mustafa, and Emad M. Alssaedi

Mobile 5G: Towards Technical View, Challenges and Issues 357
Ahmed Shamil Mustafa, Mustafa Maad Hamdi, and Mohammed Salah Abood

Federated Learning Approach for Privacy-Preserving on the D2D Communication in IoT 369
Tanweer Alam

A Performance Evaluation of Voice Over IP Protocols (SIP and H.323) in Wireless Network 381
Mustafa Sabah Noori, Ratna Kalos Zakiah Sahbudin, Mohammed Salah Abood, and Mustafa Maad Hamdi

Bringing Blockchain Technology in Innovating Industries: A Systematic Review 391
Daimler Benz, Muzaffar Hamzah, Mohd Fahmi Ghazali, and Mohammad Fadhli Asli

Improved Group-Based Signaling Scheme for Mobile Nodes in the Internet of Things Networks 417
Safwan M. Ghaleb, Mustafa Man, Baraq Ghaleb, Faizah Aplop, Mukhtar Ghaleb, and Ismail Ahmed Al-Qasem Al-Hadi

Design and Simulation a Patch Antenna for Dual-Band Frequencies . . . 435
Mohammed Salah Abood, Ahmed Shamil Mustafa, Maha Fathy, Mustafa Maad Hamdi, and Mustafa Sabah Noori

Centroidal Polygon: A New Modified Euler to Improve Accuracy of Resistor-Inductor (RL) Circuit Equation 445
S. Nooraida, Z. Nur Shahirah, S. Noor Suhana, and N. Akhyari

Investigating the Relationship Between Medicine Student's Lifestyle and Their Academic Performance 455
Yahya Almurtadha, Osama S. Mohammed, Hyder Mirghani, and Mukhtar Ghaleb

Improved Search Pattern with Discrete Wavelet Transform for Video Steganography 465
Hayfaa Abdulzahra Atee, Abidulkarim K. I. Yasari, and Dalal Abdulmohsin Hammood

Lightweight Cryptographic and Artificial Intelligence Models for Anti-smishing 483
Heider A. M. Wahsheh and Mohammed S. Al-Zahrani

URL Links Malicious Classification Towards Autonomous Threat Detection Systems 497
Mohammad Alsmadi, Izzat Alsmadi, and Heider A. M. Wahsheh

Face Recognition Using Neural Network Technique and Binary Pattern Methods 507
Mousa K. Walli, Ali Mohammed Kadhim,
and Mohammed Wajeeh Hussein

Performance Enhancements of the Indoor Localization System Based on Dynamic ToA and RSS Mechanisms 521
Zahraa Abbas, Faeza A. Abed, Nazar J. Alhyani, and Mahmood F. Mosleh

An RPL Based Optimal Sensors Placement in Pipeline Monitoring WSNs 533
Isam Wadhaj, Craig Thomson, and Baraq Ghaleb

Wireless Sensor Networks (WSN) in Oil and Gas Industry: Applications, Requirements and Existing Solutions 547
Isam Wadhaj, Craig Thomson, and Baraq Ghaleb

Applying Human Behaviour Recognition in Cloud Authentication Method—A Review 565
Mohanaad Shakir

An Intrusion Detection System Based on Hybrid of Particle Swarm Optimization (PSO) and Magnetic Optimization Algorithm (MOA) 579
Siti Norwahidayah Wahab, Noor Suhana Sulaiman, Noraniah Abdul Aziz,
Nur Liyana Zakaria, and Ainal Amirah Abd Aziz

Medical Informatics

An Enhancement of Succinate Production Using a Hybrid of Bacterial Foraging Optimization Algorithm 591
Shen Yee Siow, Mohd Saberi Mohamad, Yee Wen Choon,
Muhammad Akmal Remli, and Hairudin Abdul Majid

Analysis of Outpatient Visit Pattern and Waiting Time Prediction for Selected Public Health Clinics in Selangor 603
Suhaila Zainudin and Dzulhusni Anjang Ab. Rahman

Case Study: The Implications of Knowledge Management Tools on the Process of Overcoming COVID-19 613
Fatima Almatrooshi, Sumayya Alhammadi, Said A. Salloum,
and Khaled Shaalan

Sentiment Analysis of Arabic COVID-19 Tweets 623
Dena Ahmed, Said A. Salloum, and Khaled Shaalan

Psychometric Characteristics of the Iranian Smartphone Addiction Inventory Short Form (SPAI-SF) 633
Ibrahim Arpaci, Saeed Barzegari, Hamed Mahmoodi, Elham Afrooz, and Avishan Zanganeh Ranjbar

Diabetes Mellitus Prediction and Feature Importance Score Finding Using Extreme Gradient Boosting 643
Laboni Akter and Ferdib-Al-Islam

Factors Influencing E-payment Adoption. Case Study: Pandemic COVID-19 655
Dounia Dahab and Fatine Bouqlila

Adam Optimized Deep Learning Model for Segmenting ROI Region in Medical Imaging 669
Mustafa Musa Jaber, Sura Khalil Abd, and Saif Mohammed Ali

Analyzing the Digital Pathology Images Using Optimized Generative Adversarial Deep Learning Networks 693
Sura Khalil Abd, Mustafa Musa Jaber, and Saif Mohammed Ali

Critical Analysis of Electronic Health Readiness Assessment Factors for Development of e-Health Framework in Iraq 717
Saif Mohammed Ali, M. A. Burhanuddin, Ibrahem T. Nather, Mustafa Musa Jaber, and Sura Khalil Abd

Toward Iraqi Healthcare System Framework 737
Saif Mohammed Ali, Sura Khalil Abd, Mustafa Musa Jaber, and Ali Taha Yaseen

Advances in Information Systems

Adoption of Cloud-Based E-Government Services: A Bibliometric Analysis and Research Agenda 747
Atheer Al-Haideri, Ahmad Shukri Mohd Noor, and Ismail Ahmed Al-Qasem Al-Hadi

Implementing Knowledge Management in an IT Startup: A Case Study 757
Dena Ahmed, Said A. Salloum, and Khaled Shaalan

Managing Knowledge in Social Media 767
Mohamed Abdulla, Omar Khalid, Said A. Salloum, and Khaled Shaalan

Overviews of Uncertainty: Concepts, Categories and Coping Strategies in Decision Making 783
Aisyah Ibrahim, Tuty Asmawaty Abdul Kadir, Hamdan Daniyal, and Adzhar Kamaludin

Implementing a Talent Management System in a Group of Companies—A Case Study from UAE 795
Abdulla M. Alsharhan, Said A. Salloum, and Khaled Shaalan

Adoption of Picture Archiving and Communication System (PACS) by Healthcare Professionals 807
Ibrahim Arpaci, Saeed Barzegari, and Fatemeh Askarian

Robotic Process Automation Technology Acceptance: A Case Study from Austrian Companies 815
Martin Tschandl, Vitaliy Mezhuyev, and David Möstl

A Comprehensive Review of Innovation Adoption Types and Their Effects on the Project Performance 827
Majid G. Al-Nabae and Dania Sammani

The Role of Digital Platforms in Alleviating the Fear of Entrepreneurial Failure: A New Conceptual Framework 837
Khalid Mady, Mohamed Elsotouhy, and Mohamed Battour

The Soft Side of Technology Adoption: A Study of Customer Experience on Mobile Phone Usage 849
Yet Mee Lim, Tat-Huei Cham, Jason Wai Chow Lee, and Chuen Kee Pek

Artificial Intelligence Applications in Halal Tourism to Assist Muslim Tourist Journey 861
Mohamed Battour, Khalid Mady, Mohamed Elsotouhy, Mohamed Salaheldeen, Israa Elbendary, Mohamed Marie, and Idris Elhabony

Assistive Technology and Its Impact on Educational Achievement for Visually Impaired Students at SKPK Princess Elizabeth 873
Hussein Mohammed Esmail Abualrejal, Hassan Olayan Shtawi, Mohamad Ghozali Hassan, Amer Zaid Alqudah, and Amal Abdulwahab Alamrani

University Parcel Centre Services Quality and Users' Satisfaction in Higher Education Institutions: A Case of Universiti Utara Malaysia 885
Hussein Mohammed Abualrejal, Amer Zaid Alqudah, Ahmed Ali Atieh Ali, Oussama Saoula, and Tareq Khaled AlOrmuza

Does the Capability of Information Management, Knowledge Management, Innovation and Technology Affect the Sustainable Competitive Advantage of SMEs in Malaysia? 897
Nurul Huda Md Yatim, Najatul Afiqah Mohd Affandi, and Hazila Timan

Four-Port Wideband Circular Polarized MIMO Antenna for Sub-6 GHz Band 909
Alyaa S. Abdullah, Seham A. Hashem, Wisam S. Al-Dayyeni, Mahmood F. Mosleh, and Ihab Hassoun

Cryptographic Pseudo-Random Bit Generator Based on New Combination Discrete Chaotic Systems 921
Muhammad Hamid, Musheer Ahmad, Hussam S. Alhadawi, and Shivam Chandhok

E-Wallet Awareness: The Expansion of Knowledge in Rural Area in Terengganu Malaysia Due to Coronavirus 931
Nurul Izzah, Mazni Mohamed, Farah Dilaila, Siti Nur Fathiah, and Fify Syazwana Amelda

MySimB': An Interactive Learning Malay Idioms with Augmented Reality 945
Nur Sukinah Aziz, Azliza Yacob, Saadiah A. Rahman, Nur Liyana Zakaria, Noor Suhana Sulaiman, and Wan Nur Idayu Tun Mohd Hassan

A Review of Cyber Security Awareness (CSA) Among Young Generation: Issue and Countermeasure 957
Noor Suhana Sulaiman, Azliza Yacob, Nur Sukinah Aziz, Nooraida Samsudin, Wan Ainul Alyani Wan Mohamed, Saadiah A. Rahman, Wan Nur Idayu Tun Mohd Hassan, Akhyari Nasir, Syahrul Fahmy Abdul Wahab, and Wan Roslina Wan Othman

Extending UTAUT to Understand the Acceptance of Queue Management Technology by Physicians in UAE 969
Adi A. AlQudah and Khaled Shaalan

The Mediating Effect of Competitive Advantage on the Relationship Between Human IT Resources and Performance 983
Ehsan Ali Salem Alqararah and Wan Mohd Nazri Bin Wan Daud

A Comprehensive Analytical Case Study on Plisse Fashion Design L.L.C 993
Mohamed Abdulla, Omar Khalid, Said A. Salloum, and Khaled Shaalan

Applying Knowledge Map System for Sharing Knowledge in an Organization 1007
Fatema Hamad Mezahem, Said A. Salloum, and Khaled Shaalan

Use of Multi Agent Knowledge Management System in Technology Service Providers 1019
Fatma Ahmad Bazargan, Said A. Salloum, and Khaled Shaalan

Impact of Knowledge Management on Organizational Performance 1035
Reem Bayari, Arwa A. Al Shamsi, Said A. Salloum, and Khaled Shaalan

Artificial Intelligence and Soft Computing

ANN-Based Reference Evapotranspiration Estimation: Effects of Data Normalization and Parameters Selection

Min Yan Chia, Yuk Feng Huang, and Chai Hoon Koo

Abstract Accurate prediction of evapotranspiration (ET) is important in regions where agricultural plantations such as oil palm are abundant. This study was conducted in Peninsular Malaysia which has a large coverage of oil palm plantations that depend on rain-fed irrigation. This study attempts to improvise the estimation of reference evapotranspiration (ET_0) to aid formulation of irrigation strategies. In order to obtain desirable estimations, data pre-treatment such as normalization and input selections are essential crucial steps that are needed. Therefore, it is an aim of this study to present the effect of normalization techniques on three specific ANN-based models for estimating ET_0; namely the multilayer perceptron (MLP), radial basis function (RBF) and generalized regression neural network (GRNN). The case of different combinations of climatic parameters as input would be considered. Among the ANN models, the GRNN had the highest stability that could produce relatively stable performance regardless of the input combinations. Incorporation of normalization techniques prior to the training of the ANN-based models enabled diluting the effect of reduced input climatic parameters. For the MLP, the effect of normalization was minimal and insignificant. Selection of normalization technique of the RBF model should take the spread value of the model into consideration.

Keywords Machine learning · Normalization technique · ANN model · Reference evapotranspiration · Climatic parameter

1 Introduction

Knowledge on evapotranspiration (ET) plays an imperative role in ensuring effective and efficient water resources management [1]. Accurate and precise prediction of ET can be achieved through several methods. Conventionally, ET_0 can be

M. Y. Chia · Y. F. Huang (✉) · C. H. Koo
Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kampar, Malaysia
e-mail: huangyf@utar.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_1

calculated using the Penman–Monteith (PM) model as recommended by the Food and Agriculture Organization of the United Nations (FAO). However, this poses serious challenges to some under-developed countries where the facilities and infrastructures are insufficient or absent, to obtain all the required data.

In recent years, machine learning methods are introduced as a strategy to replace empirical methods [2]. In particular, the artificial neural network (ANN) is regarded as one of the most popular tools for ET_0 estimation [3]. ANN can be further sub-categorized into several different variations such as multilayer perceptron (MLP), radial basis function (RBF) and generalized regression neural network (GRNN). Kumar et al. [4] pioneered the transition by comparing the MLP with different architectures to lysimeter readings. The work of Kumar et al. [4] raised the interest of researchers in using the MLP for ET_0 estimation [5–7]. The RBF, which is another variation of ANN uses the Gaussian activation function in each and every hidden node in the ANN. Although the RBF has comparable performance with empirical models, however the constrained architecture of the RBF resulted in the limitation of its usage. Hence, the GRNN was introduced to solve the problem. The GRNN combines the use of the Gaussian activation of the RBF with the multiple hidden layers of MLP. The clusters and pattern trained in the pattern layer using the Gaussian activation function will be passed on to a summation layer before the output is computed. The performance of the RBF and the GRNN in estimating ET_0 was compared by Ladlani et al. [8].

The use of machine learning models to estimate ET_0 is highly data dependent. Hence, careful treatments of data are of paramount importance in order to achieve better outcome. Data pre-treatment exists in various ways. It is widely believed that normalization help to reduce the training difficulty and nullify the influence the effect of absolute scale, which may be huge when the numerical range is large [9, 10]. However, the selection of suitable normalization is deemed to be a difficult task [11–13]. The authors concluded that normalization indeed improved the accuracy but however, the improvement was not significant.

Although it has been widely accepted that normalization plays an essential role in the process of training a machine learning model, yet the investigation of the effect of the normalization techniques on the estimation of ET_0 is found to be absent unlike in other fields [12, 14]. Therefore, this study aims to examine the influence of normalization methods on ET_0 estimation using three different ANN models, which are the MLP, RBF and GRNN. The effects of different normalization methods on different ANN models are of interest in order to learn of their importance as well as compatibilities. Specifically, the z-score normalization will be compared with min–max normalization with different scales. The z-score normalization has been acclaimed to be the most commonly used normalization technique in training machine learning models [15, 16], whereas the min–max normalization is being viewed as a standard procedure in ET_0 estimation [17–19].

2 Materials and Study Area

Daily climatic parameters from 2014 to 2018 at two stations in Peninsular Malaysia, namely the KLIA Sepang (2°44′ N, 101°42′ E) and the Sitiawan (4°13′ N, 100°42′ E) weather stations were obtained from the Malaysian Meteorological Department (MMD). The climatic parameters included maximum temperature (T_{max}, °C), minimum temperature (T_{min}, °C), mean temperature (T_{mean}, °C), relative humidity (RH, %), global radiation (R_s, MJm^{-2}) and wind speed (u, ms^{-1}). ET_0 estimation for the above two stations are important as the areas are mainly covered with oil palm plantation which is virtually depending on only rain-fed irrigation, as in almost all of the plantations in the country. Both stations are at proximity to the equator of Earth. They have tropical climate which is hot and humid throughout the year. The locations of the stations are depicted in Fig. 1.

Different combinations of climatic parameters would affect the performance of the ANN models. The selected combinations (comprised of different number of climatic parameters) are summarized in Table 1. The parameters were included into the model one at a time based on their correlation with the ET_0, where higher correlation would come first.

3 Methods

3.1 FAO-56 Penman Monteith Equation

Prior to the training of the ANN models, a set of standard ET_0 should be calculated in order to perform the supervised learning. The PM model was used to calculate the actual ET_0 at the KLIA Sepang and Sitiawan stations. This model was selected as the reference model due to the endorsement of the FAO and many other literatures [20–22]. The equation of PM model is shown in Eq. 1.

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma\left(\frac{900}{T + 273}\right)u_2(e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (1)$$

where ET_0 is the daily reference evapotranspiration (mm/day), R_n is the net radiation (MJ m^{-2} day^{-1}), G is the soil heat flux (MJ m^{-2} day^{-1}), T is the daily mean temperature at 2 m height (°C), u_2 is the wind speed at 2 m height (m/s), e_s is the mean saturation vapor pressure (kPa), e_a is the actual vapour pressure (kPa), Δ is the slope of vapour pressure curve (kPa/°C) and γ is the psychrometric constant (kPa/°C).

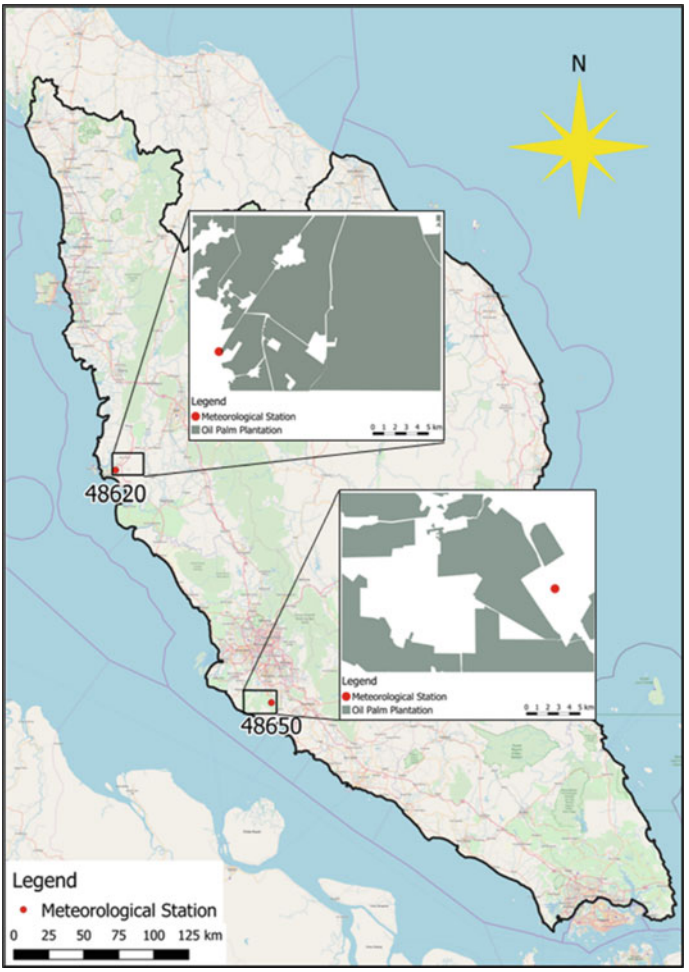


Fig. 1 Locations of the investigated stations

Table 1 Summary of different combinations of climatic parameters used

Combination	Climatic parameters
C1	R_s
C2	R_s, u
C3	T_{max}, R_s, u
C4	T_{max}, RH, R_s, u
C5	T_{max}, T_{mean}, RH, R_s, u
C6	T_{max}, T_{min}, T_{mean}, RH, R_s, u

3.2 ANN Models

Three ANN models, namely the MLP, RBF and GRNN models were used in this study. In the MLP, inputs were fed into a network with one or more hidden layers and varying number of nodes. In this study, the MLP with one hidden layer was used. In the hidden layer, the number of hidden nodes would be one extra as compared to the input nodes. As for the RBF model, only three layers (input layer, hidden layer and output layer) are present. The activation function used in the hidden layer is a Gaussian function. Lastly, the GRNN has four layers, with the inclusion of a pattern layer and a summation layer for training and computing purposes. In the pattern layer, Gaussian function was also used as the activation function. All the architectures of the ANN models were selected based on a series of trial and error.

3.3 Normalization Techniques

The effect of three different normalization techniques on the ANN models were investigated in this study. The first method is the conventional z-score normalization which transforms the original data into a set of data with mean at 0 and standard deviation of 1. The equation for z-score normalization in shown in Eq. 2.

$$x_{norm} = \frac{x - \mu}{\sigma} \tag{2}$$

where x is the original data, x_{norm} is the normalized data, μ is the mean of data and σ is the standard deviation of data.

In the case of the min–max normalization, the original data were rescaled to range between 0 and 1 (or −1 and 1) based on their numerical value in absolute scale. The equation to perform such normalization is as Eq. 3.

$$x_{norm} = \frac{x_0 - x_{min}}{x_{max} - x_{min}} \tag{3}$$

where x_{max} is the maximum value of data and x_{min} is the minimum value of data.

For the min–max normalization which range between −1 and 1, Eq. 3 can be modified into Eq. 4.

$$x_{norm} = 2\left(\frac{x_0 - x_{min}}{x_{max} - x_{min}}\right) - 1 \tag{4}$$

3.4 Performance Indicators

In order to evaluate the performances of the ANN models, several performance measure metrics were used. The equations to calculate the mean absolute error (MAE), root mean square error (RMSE) and Nash–Sutcliffe efficiency (NSE) are shown in Eqs. 5–7, respectively.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i}^{N}\left|y_{actual,i} - y_{predicted,i}\right| \tag{5}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i}^{N}\left(y_{actual,i} - y_{predicted,i}\right)^2} \tag{6}$$

$$\mathrm{NSE} = 1 - \frac{\sum_{i}^{N}\left(y_{actual,i} - y_{predicted,i}\right)^2}{\sum_{i}^{N}\left(y_{actual,i} - y_{mean}\right)^2} \tag{7}$$

where $y_{actual,i}$ is the ith actual ET_0 value and $y_{predicted,i}$ is the ith predicted ET_0 value by ANN model.

4 Results and Discussion

Six combinations (represented by C1–C6) were tested to examine how with the different number of climatic parameters fed into the machine learning models would affect their performances. The overall results at Station 48650 (KLIA Sepang) are arbitrarily selected for discussion and are summarized in Fig. 2. It can be seen that the MLP was sensitive to the number of climatic parameters, which was revealed by the relative low NSE value and high MAE as well as RMSE when C1 was used. The performance of the model started to improve when more climatic parameters were used and the model performed better from C2–C6. Normalization enhanced the performance of the model but could not remove the effect of input climatic parameters.

For the RBF, there was no clear relationship between the model performance and the number of input parameters. This could be due to the absence of normalization which allowed the impact of absolute numerical value dominating the training process [9]. After performing normalization, C2 and C3 had the highest MAE and RMSE, with C3 reporting the lowest NSE value as well. Unlike the MLP and RBF, the GRNN had the highest model stability towards the number of input climatic parameters where all input combinations achieving similar MAE, RMSE and NSE. In other words, this suggests that the GRNN ought to be used if the area has limited available meteorological data for ET_0 estimation in order to obtain stable and highly reliable results.

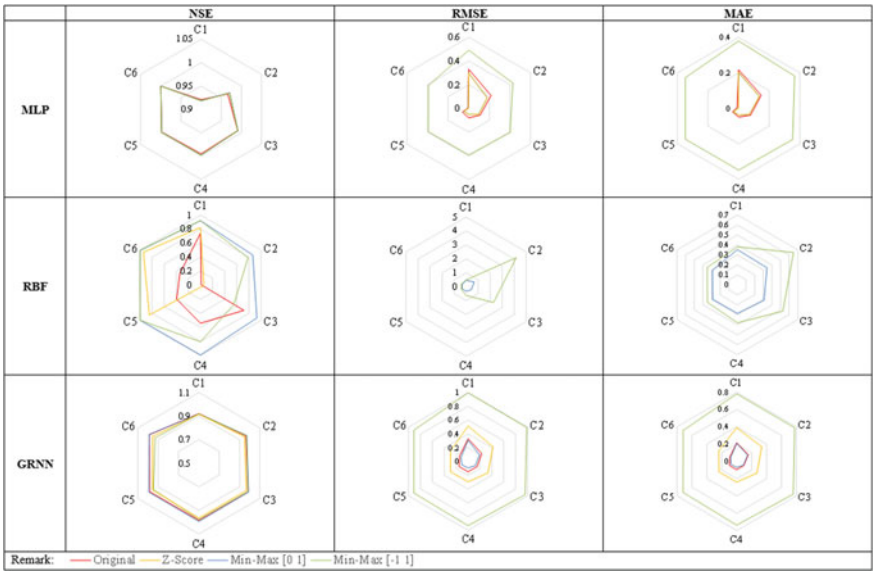


Fig. 2 Radar charts of performance metrics for station 48650 (KLIA Sepang) using different machine learning models and normalization techniques (RMSE and MAE for "original" and "z-score" normalisation of RBF are not presented due to the extremely large values which are not compatible with the scale used)

The ability of the MLP to handle highly non-linear problems had diluted the effect of normalization techniques in the machine learning training process. However, it was worthy to note that changing the range of min–max normalization would impair the performance of the MLP model as a whole. When the range of min–max normalization increases, the outliers were being pulled away from the central tendencies. This case worsened with the use of −1 and 1 as the beginning and end of the intervals. Using negative values would have certain effect during the denormalization process, which would influence the accuracy of ET_0 estimation. For the MLP, the z-score and min–max [0 1] normalizations are recommended as both of the normalization techniques were able to improve the estimation MLP even though original data estimation had already produced outstanding results.

In contrast with the MLP, the RBF based machine learning models generally incurred large errors with low model stability (as indicated by the low NSE values) when estimating ET_0 using original data. This suggested that the presence of normalization techniques is critical for training an RBF model. Interestingly, unlike in the case of MLP, the z-score normalization could not produce a positive impact on the accuracy of RBF model. In fact, for C2, C3 and C4, the performance of the RBF trained with z-score normalized data were actually worse than the original data. It can be inferred that the performance of RBF is highly dependent on the designed spread size [23]. In this study, the spread size used for the RBF model was 1 for all cases. Consequently, the radius for the RBF model could be too small to

signify the importance of the distances between data points of original data or data normalized using z-score normalization. The role of min–max [−1 1] normalization on stabilizing the performance of the RBF model was revealed obviously by how the normalization reduced the MAE and RMSE while providing consistent NSE. Furthermore, the min–max [−1 1] normalization appeared to reduce the effect of different input combinations except for C2 and C3 which had slightly higher MAE and RMSE. For the case of C1, C4, C5 and C6, the errors and stability of the RBF model predictions were similar, which proved the importance of min–max normalization in improving model performance. The min–max [0 1] normalization can be regarded as the most suitable normalization method for the RBF models. The RBF models trained using min–max [0 1] normalized data had the lowest reported errors as well as highest NSE values.

Regardless of the normalization methods used, the performance of the GRNN was consistent for all input combinations. That is to say, the stability of GRNN models is higher when compared to the MLP and RBF, for the case of limited climatic parameters. Nevertheless, the effect of normalization methods on the GRNN should be addressed. As shown in Fig. 3, the min–max [0 1] normalization enhanced the performance of the model whereas the z-score as well as min–max [−1 1] normalization method reduced the accuracy of the GRNN predictions (Fig. 3).

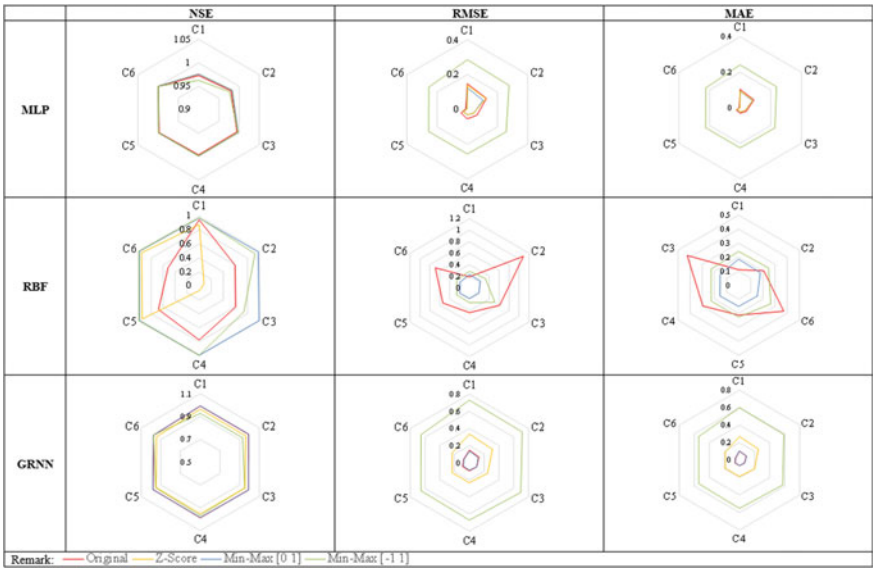


Fig. 3 Radar charts of performance metrics for station 48620 (Sitiawan) using different machine learning models and normalization techniques (RMSE and MAE for "original" and "z-score" normalisation of RBF are not presented due to the extremely large values which are not compatible with the scale used)

5 Conclusion

This paper discusses the potential of three machine learning models, namely the MLP, the RBF and the GRNN for estimating ET_0 with different input parameters in a tropical climate region where water management for oil palm plantations are crucial. It was found that the GRNN models were the most stable models towards different input combinations followed by the MLP and then the RBF. Normalization could help to dilute the effect of different input combinations. The MLP which can handle highly non-linear problems performed well even without normalization. For the RBF, selection of normalization techniques should take into consideration the spread value of the RBF network. In this study, the min–max [0 1] was the most suitable normalization when the spread value used was 1. Lastly, the effect of the normalization techniques on the GRNN could be viewed as the combined impact of them, on both the MLP and the RBF taken separately. Similar patterns and findings were discovered at both stations considered in this study. From the geographical position of the two stations along the west coast of Peninsular Malaysia, it could be concluded that the behaviors of machine learning models and normalization techniques should be similar along the entire coast line. The GRNN coupled with the min–max [0 1] normalization should be considered first for its stable performance as well as the accurate prediction of ET_0 for the stated region. This work was restricted in both spatial and temporal dimension due to the lack of availability of data, and thus the model could be further proven when more data can be accessed in the future.

Acknowledgements This research was funded by Universiti Tunku Abdul Rahman (UTAR), Malaysia through the Universiti Tunku Abdul Rahman Research Fund under project number IPSR/RMC/UTARRF/2018-C2/K03.

References

1. da Silva Júnior JC, Medeiros V, Garrozi C, Montenegro A, Gonçalves GE (2019) Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. Comput Electr Agric 166:105017
2. Chia MY, Huang YF, Koo CH, Fung KF (2020) Recent advances in evapotranspiration estimation using artificial intelligence approaches with a focus on hybridization techniques—a review. Agronomy 10(1):101
3. Adamala S (2019) Nonlinear evapotranspiration modeling using artificial neural networks. Adv Evapotranspiration Methods Appl
4. Kumar M, Raghuwanshi NS, Singh R, Wallender WW, Pruitt WO (2002) Estimating evapotranspiration using artificial neural network. J Irrig Drain Eng 128(4):224–233
5. Reis MM, da Silva AJ, Zullo Junior J, Tuffi Santos LD, Azevedo AM, Lopes ÉMG (2019) Empirical and learning machine approaches to estimating reference evapotranspiration based on temperature data. Comput Electr Agric 165:104937

6. Antonopoulos VZ, Antonopoulos AV (2017) Daily reference evapotranspiration estimates by artificial neural networks technique and empirical equations using limited input climate variables. Comput Electron Agric 132:86–96
7. Pandey PK, Nyori T, Pandey V (2017) Estimation of reference evapotranspiration using data driven techniques under limited data conditions. Model Earth Syst Environ 3(4):1449–1461
8. Ladlani I, Houichi L, Djemili L, Heddam S, Belouz K (2012) Modeling daily reference evapotranspiration (ET0) in the north of algeria using generalized regression neural networks (GRNN) and radial basis function neural networks (RBFNN): a comparative study. Meteorol Atmos Phys 118(3–4):163–178
9. Feng Y, Peng Y, Cui N, Gong D, Zhang K (2017) Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data. Comput Electron Agric 136:71–78
10. Nourani V, Elkiran G, Abdullahi J (2019) Multi-station artificial intelligence based ensemble modeling of reference evapotranspiration using pan evaporation measurements. J Hydrol 577:123958
11. Jain S, Shukla S, Wadhvani R (2018) Dynamic selection of normalization techniques using data complexity measures. Expert Syst Appl 106:252–262
12. Singh D, Singh B (2019) Investigating the impact of data normalization on classification performance. Appl Soft Comput 105524
13. Pan J, Zhuang Y, Fong S (2016) The impact of data normalization on stock market prediction: using SVM and technical indicators. In: Soft computing in data science. Communications in computer and information science, vol 652, pp 72–88
14. Shahriyari L (2019) Effect of normalization methods on the performance of supervised learning algorithms applied to HTSeq-FPKM-UQ data sets: 7SK RNA expression as a predictor of survival in patients with colon adenocarcinoma. Brief Bioinform 20(3):985–994
15. Kotu V, Deshpande B (2019) Classification, data science: concepts and practice. Morgan Kaufmann
16. Yamaç SS, Todorovic M (2020) Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data. Agric Water Manage 228:105875
17. Dou X, Yang Y (2018) Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems. Comput Electron Agric 148:95–106
18. Wu L, Zhou H, Ma X, Fan J, Zhang F (2019) Daily reference evapotranspiration prediction based on hybridized extreme learning machine model with bio-inspired optimization algorithms: application in contrasting climates of China. J Hydrol 577:123960
19. Huang G, Wu L, Ma X, Zhang W, Fan J, Yu X et al (2019) Evaluation of CatBoost method for prediction of reference evapotranspiration in humid regions. J Hydrol 574:1029–1041
20. Allen RG, Pereira L, Raes D, Smith M (1998) Crop evapotranspiration—guidelines for computing crop water requirements. FAO Irrigation and Drainage Paper 561998
21. Chia MY, Huang YF, Koo CH (2021) Swarm-based optimization as stochastic training strategy for estimation of reference evapotranspiration using extreme learning machine. Agric Water Manage 243:106447
22. Chia MY, Huang YF, Koo CH (2020) Support vector machine enhanced empirical reference evapotranspiration estimation with limited meteorological parameters. Comput Electron Agric 175:105577
23. Ghosh-Dastidar S, Adeli H, Dadmehr N (2008) Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. IEEE Trans Biomed Eng 55(2 Pt 1):512–518

Trend Analysis of the Effects of Climate Change on the Pan Evaporation Rate in Sabah, Malaysia

Min Yan Chia, Yuk Feng Huang, Chai Hoon Koo, and Yeu Zhi Tan

Abstract The pan evaporation rate is an important measurement that can be used to derive other meteorological parameter such as the evapotranspiration and drought indices. Learning the trend of pan evaporation rate is useful to project the possible trends of other important parameters. In this study, trend analysis of pan evaporation rate was conducted at four stations, namely at Station 96471 (Kota Kinabalu), Station 96477 (Kudat), Station 96481 (Tawau) and at Station 96491 (Sandakan), to evaluate the effect of climate change on the pattern of pan evaporation rate from January 1985 to December 2019. The Mann Kendall (MK) test (assisted by the Sen's slope estimator) and the Holt's Linear Trend Method (HLTM) were used to perform the trend analysis. It was found that the pan evaporation rate was sensitive to climatic conditions such as drought and rainfall. Generally, the trend reported in the MK test and Sen's Slope estimator tallied with that of the HLTM. However, in some instances, the HLTM was less sensitive as it could not detect the trends that were alerted by the MK test and Sen's slope estimator. Overall, the findings of this study was in agreement with other researchers conducted in similar regions, and therefore could be a useful tool in the decision making process.

Keywords Trend analysis · Mann Kendall · Sen's slope estimator · Holt's linear trend method · Pan evaporation · Climate change

1 Introduction

Global climate change poses immense challenge on the water management of every nation. The report from the Intergovernmental Panel on Climate Change (IPCC) indicated that the increase in the world's average temperature had doubled over the millennium century as compared to the pre-industrial era [1]. Consequently, there

M. Y. Chia · Y. F. Huang (✉) · C. H. Koo · Y. Z. Tan
Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kampar, Malaysia
e-mail: huangyf@utar.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_2

had been changes in the natural phenomena such as rainfall, occurrence of flash floods and droughts [2]. Evaporation is one of the contributors of the loss of surface water from the Earth alongside with transpiration. In a typical tropical rainforest, evaporation accounts for about 40% of water loss after correction for possible runoffs [3]. As compared to obtaining the values of transpiration, the measurement of evaporation is far easier using the pan method, while many other empirical models could carry out similar function [4]. On top of that, the pan evaporation is an important measurement used to calculate other meteorological parameters such as the evapotranspiration and various drought indices [5]. This has raised the interest of researchers to investigate the pattern of the evaporation processes.

There have been previous studies that had performed trend analysis of evaporation in various regions with different climate patterns. Singh, et al. examined the trend of pan evaporation in India using the Mann Kendall (MK) test as well as testing the anomalies [6]. It was found that at the monthly resolution, the pan evaporation exhibited decreasing trend, where no shift was detected after the Pettitt-Mann–Whitney test was performed. Numerous studies also opted for the non-parametric MK test to observe the temporal variability of pan evaporation over a selected period, with some adding the Thiel-Sen estimator to determine the momentum of trends [7, 8]. In the case where serial correlation has to be removed, the MK test was improvised by performing a pre-whitening treatment to the data series [9]. The literature review reveals that the MK test had been a dominant method in assessing the pan evaporation in various regions, while suitability of other methods yet to be ascertained. Moreover, the trends of pan evaporation differ from each other due to the diversity of climate as well as land use and this would require more initiative to conduct robust trend analysis wherever necessary [10, 11].

In this paper, the authors are interested in investigating the effect of climate change on the trend of pan evaporation in the Sabah state, which is located in the East Malaysia. Sabah has typical tropical climate and is suitable for the oil palm plantation, which almost depends only on rain-fed irrigation. The trend analysis was done using two approaches, namely Mann Kendall (MK) test combined with the Sen's slope estimator, and the Holt's Linear Trend Method (HLTM). Trend analysis has yet to be conducted in tropical regions like Sabah, Malaysia, let alone the subject matter of this study focusing on oil palm plantations. On top of that, the utilization of the HLTM as a tool for trend analysis has not been compared with generally accepted approaches like the MK test and the Sen's slope estimator. Therefore, the attributes and characteristics of HLTM in performing trend analysis is another important aspect of this research work.

2 Materials and Study Area

Monthly pan evaporation data from January 1985 to December 2019 (420 months in 35 years) were collected from the Malaysian Meteorological Department (MMD). These data were inclusive for four meteorological stations in the Sabah

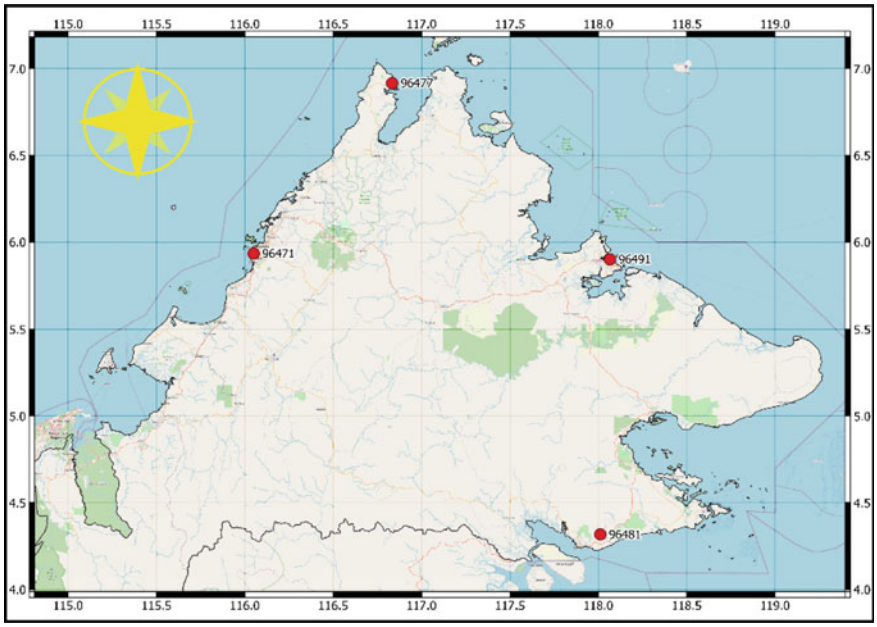


Fig. 1 Locations of the stations investigated

state, namely the Station 96471 (Kota Kinabalu), Station 96477 (Kudat), Station 96481 (Tawau) and the Station 96491 (Sandakan). Sabah was selected as the study area due to its high coverage of oil palm plantations which is crucial to support the nation's gross domestic production (GDP). The pan evaporation rate significantly affects the water available for the agricultural crop and hence the monitoring of its trend is imperative. The locations of the stations are shown in Fig. 1.

According to a previous study, regions around Station 96471 (Kota Kinabalu) and Station 96491 (Sandakan) have high risk of water insecurity. On the contrary, Kudat is less vulnerable to issues related to water supply [12]. There is no available data for Station 96481 (Tawau) and its nearby area. Hence, the selection of study areas should provide interesting findings based on their natures and distributions.

3 Methods

3.1 Mann Kendall Test

The MK test is a non-parametric test. It is a useful trend test to provide trend statistics in many meteorological parameters, including the pan evaporation rate. This test is commonly used as it does not require any assumptions. Therefore, the

MK test is very suitable for data which are not normally distributed. Equation 1 shows the calculation of the MK test statistics.

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \operatorname{sgn}(x_j - x_i) \tag{1}$$

where S is the MK test statistics, x are the observations and sgn() is a sign function to dictate the sign of $(x_j - x_i)$ as shown in Eq. 2.

$$\operatorname{sgn}(x_j - x_i) = \begin{cases} +1 & \text{if}\left(x_j - x_i\right) > 0 \\ 0 & \text{if}\left(x_j - x_i\right) = 0 \\ -1 & \text{if}\left(x_j - x_i\right) < 0 \end{cases} \tag{2}$$

Subsequently, the variance, V(S) as well as the normalised value of S, Z are computed using Eqs. 3 and 4, respectively.

$$V(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i - 1)(2t_i + 5)}{18} \tag{3}$$

$$Z = \begin{cases} \frac{S-1}{\sqrt{V(S)}} & \text{if } S > 0 \\ \frac{S+1}{\sqrt{V(S)}} & \text{if } S < 0 \end{cases} \tag{4}$$

where m is the number of tie groups and t is the size of the tie. The sign of Z value determines whether the pan evaporation rate is on an increasing or decreasing trend. The significance of the trend is decided based on a significant level at 5%. Some research papers suggested that pre-whitening should be done prior to the trend analysis. However, in this study, the pre-whitening stage can be omitted due to the low variation of the data set, which fell between 0.1 and 0.2 [13].

3.2 *Sen's Slope Estimator*

The Sen's slope estimator is another non-parametric test named after Theil-Sen. Conventionally, the Sen's slope estimator is used as a complementary tool to estimate the true value after the analyzing the trends using the MK test (provided that they have a significant trend). The Sen's slope estimates (Q_β) are less sensitive to the outliers and can be computed using Eq. 5.

$$Q_\beta = \text{median}\left(\frac{x_j - x_i}{j - i}\right) \tag{5}$$

where x are the observations and j must be greater than i.

3.3 Holt's Linear Trend Method

The Holt's Linear Trend Method (HLTM) is also known as the exponentially-weighted moving average method. This exponential smoothing is useful to observe the seasonality or change in trend direction by taking the past observations in to account. In this study, the smoothing constant, α was set to be equal to 0.1. The equations of the Holt's Linear Trend Method can be expressed as shown in Eq. 6.

$$S_t = \alpha y_{t-1} + (1 - \alpha) S_{t-1} \tag{6}$$

where S is the value of the smoothened series and y is the value of the original series.

3.4 Overall Workflow

The trend analysis of pan evaporation rate at the four selected stations were done at two scales. In the macro scale, the 35 years of data with 420 months of data points were analysed as a whole, providing the trend pattern of pan evaporation rate for the whole time span of 35 years. At the other scale, the whole 35 years duration was subdivided in to seven sub-windows, whereby each sub-window consists of five years of data. In this way, the micro scale of the pan evaporation rate at the four stations can be observed in a more detailed manner. The findings using the MK test assisted by Sen's slope estimator were compared to those of the HLTM.

4 Results and Discussion

In Fig. 2, the overall trends of pan evaporation rate from January 1985 to December 2019 for the Station 96471 (Kota Kinabalu), Station 96477 (Kudat), Station 96481 (Tawau) and the Station 96491 (Sandakan) are shown. In the macro scale, there is no significant changes in the pan evaporation rate across the 35-year time span as suggested by the smoothen series obtained through the HLTM. Nevertheless, some common patterns and attributes can be extracted from the Holt's smoothened series.

As shown in Fig. 2, there was a considerable drop in the pan evaporation rate around the year 1998–2001 ay all the four studied stations. This finding is interesting whereby previous studies claimed that the aforementioned period had the relatively lower drought intensity as well as higher rainfall [14, 15]. It can be inferred that the high ambient humidity from 1998 to 2001 had contributed to the lower pan evaporation rate in the area. This also implied that the pan evaporation is sensitive to the change in the climatic condition, which can be more drastic when

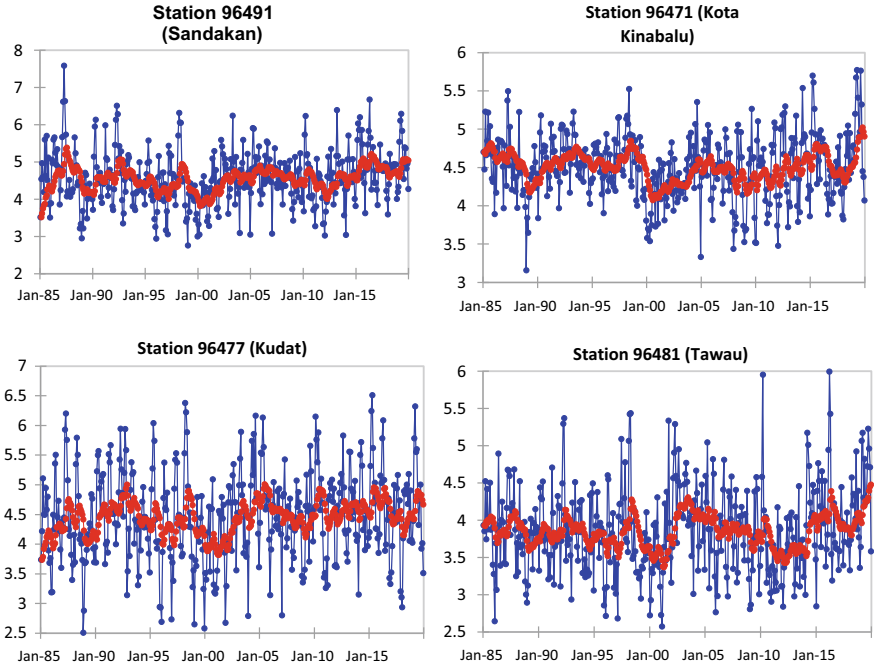


Fig. 2 Overall trend of pan evaporation rate with HLTM smoothen series at the studied stations

the climate change becomes more and more prominent. In order to support the statement, the authors examined the pan evaporation rate pattern during the occurrence of the El-Nino drought during 1997–1998. The rainfall measured during this period had significantly decreased thus leading to a drier environment [16]. Consequently, the pan evaporation rate of the four stations during this drought stretch had increased as depicted in Fig. 2. It is important for the policy makers to realised and understand the relationship of the gradual or sudden change in the climate pattern to the trend of the pan evaporation rate as the mortality of low land tree species had increased substantially through the El-Nino drought.

Table 1 shows the trend analysis pattern evaluated by the MK test assisted by the Sen's slope estimator. The signs of the normalised MK test statistics indicate whether the trend is increasing (positive) or decreasing (negative). On top of that, the magnitude of the Sen's slope estimator carries the information of the intensity of increment or decrement. Periods with decreasing pattern of pan evaporation rate are shaded in grey. Figure 3 summarises the trend analysis of the pan evaporation rates at the four studied stations graphically, whereby the results using MK test can be compared and contrasted with its corresponding HLTM counterpart.

From Table 1, it can be seen that the pan evaporation rates were on the negative trend at all the stations from year 1985 to 1999. Generally, the results tabulated in Table 1 matches the smoothened trend in Fig. 3 obtained using the HLTM. However, the MK test and Sen's slope estimator are more sensitive as compared to

Table 1 Trend analysis of pan evaporation rate in different stations using MK Test and Sen's slope estimator

		Station 96471 (Kota Kinabalu)	Station 96477 (Kudat)	Station 96481 (Tawau)	Station 96491 (Sandakan)
1985–1989	Z	−3.514	−1.339	−1.435	−2.609
	SS	−0.009	−0.007	−0.004	−0.015
1990–1994	Z	−1.218	−2.583	−1.167	−2.003
	SS	−0.003	−0.013	−0.005	−0.009
1995–1999	Z	−0.810	−1.448	−0.925	−1.652
	SS	−0.002	−0.009	−0.004	−0.009
2000–2004	Z	4.031	3.259	2.621	2.417
	SS	0.012	0.018	0.013	0.013
2005–2009	Z	−1.932	−1.563	−1.244	−2.883
	SS	−0.006	−0.008	−0.005	−0.011
2010–2014	Z	0.657	−1.116	1.856	0.236
	SS	0.003	−0.007	0.008	0.001
2015–2019	Z	0.440	−0.897	1.652	−0.874
	SS	0.002	−0.005	0.008	−0.004

Z = Normalised MK test statistics, SS = Sen's slope

the HLTM. For instance, in most cases, when the trend line of the HLTM suggested that there was no significant increase in the pan evaporation rate, the MK test and Sen's slope estimator, had on contrary reported otherwise. One of the obvious examples appeared in year 1990–1994 at Station 96477 (Kudat). From Table 1, the reported normalised MK test statistics and Sen's slope are −2.583 and −0.013, respectively. However, the HLTM suggested that there was no significant trend based on the plotted trend line. In fact, the trend line tilted upwards at the end of the 5-year window. Therefore, it can be said that the MK test are over-sensitive, and this may lead to the detection of the undesirable "fake trend".

Also, it was observed that the trends of pan evaporation rate at Station 96471 (Kota Kinabalu) and Station 96481 (Tawau) were similar to each other. On the other hand, Station 96477 (Kudat) and Station 96491 (Sandakan) shared like trends. This can be used by the policy makers to formulate different water management strategies based on the clusters.

Overall, the trends detected by the MK test, Sen's slope estimator and HLTM are in agreement with the findings of previous studies. The rate of warming in Malaysia became significant after the year 1971 and the effect of global warming can be delayed for at least 10 years [17, 18]. Therefore, it is reasonable that the pan evaporation rate started to exhibit drastic changes after the millennium. This also tallies with researches that were conducted in Thailand and New Zealand, with the former being located in the proximity of the Equator, whereby the pan evaporation rates of these two regions also experienced pattern reversal after the year 2000 [19, 20]. Hence, the conclusion obtained from the trend analysis can be justified accordingly.

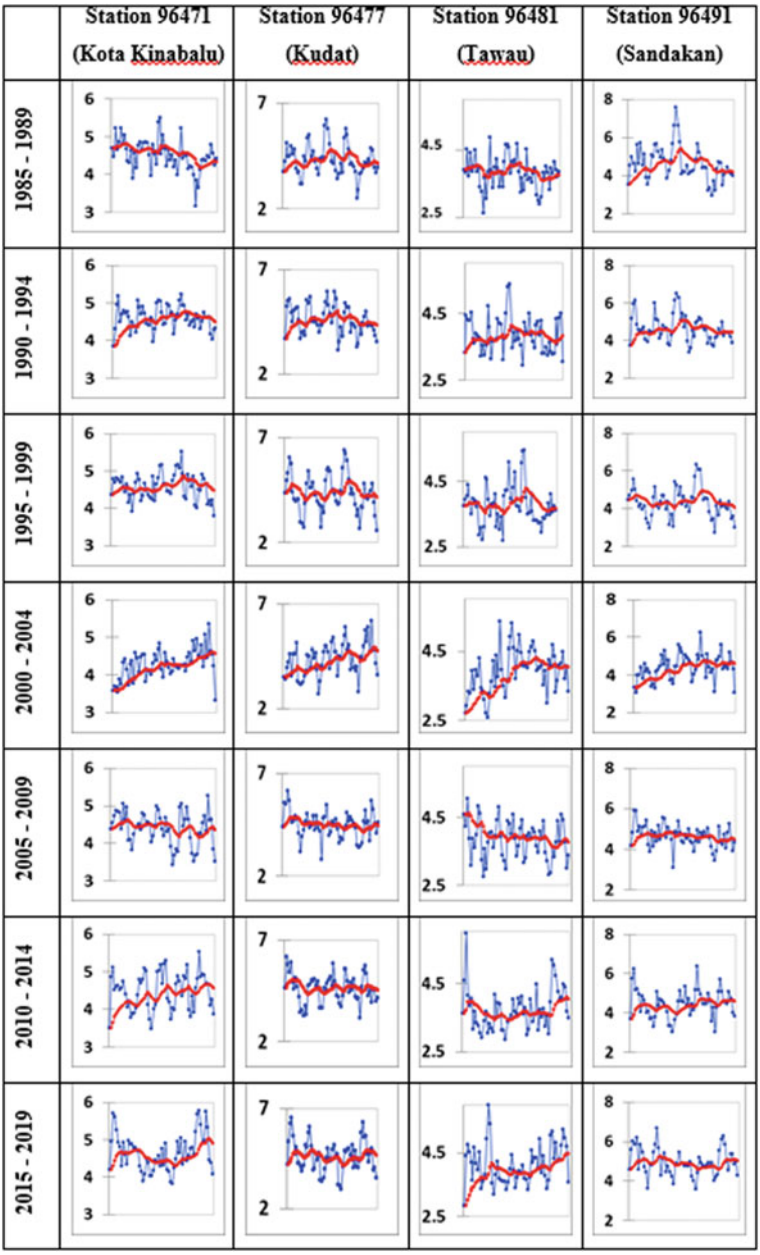


Fig. 3 HLTM smoothen series (red line) of pan evaporation rate at the studied stations

5 Conclusion

Trend analysis was conducted at four stations in the Sabah state of East Malaysia, namely at Station 96471 (Kota Kinabalu), Station 96477 (Kudat), Station 96481 (Tawau) and at Station 96491 (Sandakan) using the MK test, the Sen's slope estimator and the HLTM. It was found that the pan evaporation rate was sensitive to the change in the climatic condition, including drought (occurrence of El-Nino) and increase in rainfall. The MK test and Sen's slope estimator were reported to be more sensitive than the HLTM, whereby in some of the cases, the trend lines of HLTM could not showed the trend suggested by the former two analysis tools. In conclusion, the trend analysis performed in this study was in agreement with similar investigations conducted in other regions close to the Equator. Therefore, the findings can be used as an important tool for the policy makers to plan the allocation of water resources effectively.

Acknowledgements This research was funded by Universiti Tunku Abdul Rahman (UTAR), Malaysia through the Universiti Tunku Abdul Rahman Research Fund under project number IPSR/RMC/UTARRF/2018-C2/K03. The meteorological data were supplied by the Malaysian Meteorological Department (MMD).

References

1. Intergovernmental Panel on Climate Change (2019) Global warming of 1.5 °C
2. Amin MZM, Shaaban AJ, Ercan A, Ishida K, Kavvas ML, Chen ZQ et al (2017) Future climate change impact assessment of watershed scale hydrologic processes in Peninsular Malaysia by a regional climate model coupled with a physically-based hydrology modelo. Sci Total Environ 575:12–22
3. Kunert N, Aparecido LMT, Wolff S, Higuchi N, Santos Jd, Araujo ACd, et al (2017) A revised hydrological model for the Central Amazon: the importance of emergent canopy trees in the forest water budget. Agric Forest Meteorol 239:47–57
4. Wang L, Han S, Tian F (2021) Comparison of formulating apparent potential evaporation with pan measurements and Penman methods. J Hydrol 592:125816
5. Mahmud K, Siddik M, Khatun M, Islam M (2020) Performance evaluation of Class A pan coefficient models to estimate reference evapotranspiration in Mymensingh region of Bangladesh. J Bangladesh Agric Univ (0):1
6. Singh D, Kumar AR, Goyal V, Arora M, Allaka N (2020) Temporal characteristics of evaporation trends and their attribution to meteorological drivers in Roorkee (India). Authorea
7. Li M, Chu R, Shen S, Islam A (2018) Dynamic analysis of pan evaporation variations in the Huai River Basin, a climate transition zone in eastern China. Sci Total Environ 625:496–509
8. Vicente-Serrano SM, Bidegain M, Tomas-Burguera M, Dominguez-Castro F, El Kenawy A, McVicar TR et al (2018) A comparison of temporal variability of observed and model-based pan evaporation over Uruguay (1973–2014). Int J Climatol 38(1):337–350
9. Xu Y-P, Pan S, Gao C, Fu G, Chiang Y-M (2016) Historical pan evaporation changes in the Qiantang River Basin, East China. Int J Climatol 36(4):1928–1942
10. Mhawej M, Fadel A, Faour G (2020) Evaporation rates in a vital lake: a 34-year assessment for the Karaoun Lake. Int J Remote Sens 41(14):5321–5337

11. Hounguè R, Lawin A, Moumouni S, Afouda A (2018) Change in climate extremes and pan evaporation influencing factors over Ouémé Delta in Bénin. Climate 7(1):2
12. Payus C, Ann Huey L, Adnan F, Besse Rimba A, Mohan G, Kumar Chapagain S et al (2020) Impact of extreme drought climate on water security in North Borneo: case study of Sabah. Water 12(4):1135
13. Bayazit M, Önöz B (2007) To prewhiten or not to prewhiten in trend analysis? Hydrol Sci J 52(4):611–624
14. Boyd DS, Foody GM, Phipps PC (2007) Dynamics of ENSO drought events on Sabah rainforests observed by NOAA AVHRR. Int J Remote Sens 27(11):2197–2219
15. Newbery DM, Lingenfelder M (2004) Resistance of a lowland rain forest to increasing drought intensity in Sabah, Borneo. J Trop Ecol 20(6):613–624
16. Aiba S-I, Kitayama K (2002) Effects of the 1997–98 El Niño drought on rain forests of Mount Kinabalu, Borneo. J Trop Ecol 18(2):215–230
17. Shaffril HAM, Samah BA, D'Silva JL, Uli J (2011) Global warming at the East Coast zone of Peninsular Malaysia. Am J Agric Biol Sci 6(3):377–383
18. Budzianowski WM (2011) Time delay of global warming. Int J Glob Warming 3(3):289
19. Tebakari T, Yoshitani J, Suvanpimol C (2005) Time-space trend analysis in pan evaporation over Kingdom of Thailand. J Hydrol Eng 10(3):205–215
20. Roderick ML, Farquhar GD (2005) Changes in New Zealand pan evaporation since the 1970s. Int J Climatol 25(15):2031–2039

Cow Image Segmentation Using Mask R-CNN Integrated with Grabcut

Rotimi-Williams Bello, Ahmad Sufril Azlan Mohamed, and Abdullah Zawawi Talib

Abstract Different state-of-the-art object detection methods have been applied in agriculture for precision livestock farming. However, the quality and accuracy of livestock such as cows being detected in an image during computer vision tasks depend on the segmentation and extraction techniques used. Mask R-CNN, an instance segmentation method popular for its class and mask regression has been widely applied for cow image segmentation tasks. However, its algorithm relies on simultaneous localization and mapping algorithms, thereby discrediting its ability to completely segment an image foreground from the image background. In this paper, a Mask R-CNN segmentation method integrated with Grabcut is proposed. The method is for the detection and complete extraction of an image foreground from the image background. We performed an experiment using edge detection to compare the segmentation that is region-based with the boundary estimation. After comparing the segmentation, we have concluded that the Grabcut integrated with the Mask R-CNN is based on the computation of the edge detection approach employed by our proposed method.

Keywords Cow · Grabcut · Mask R-CNN

R.-W. Bello (✉) · A. S. A. Mohamed · A. Z. Talib
School of Computer Sciences, Universiti Sains Malaysia, 11800 Pulau Pinang, Malaysia
e-mail: sufril@usm.my

A. Z. Talib
e-mail: azht@usm.my

R.-W. Bello
Department of Mathematics/Computer Sciences, University of Africa, Toru-Orua, Sagbama, Bayelsa State, Nigeria

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_3

1 Introduction

Livestock application of computer vision and image processing algorithms involve some steps of which image segmentation is an important step. The image segmentation steps are necessary because the diversity in species of most animals can only be easily represented using breeds, and the identification of these breeds is important for precision animal husbandry. The two characterization approaches used in recognizing the breeds of animal are genotypic characterization approach and phenotypic characterization approach [1, 2]. The genotypic approach is necessary for understanding genetic resource diversity and uniqueness to frame policy which improves the value and scope of the resource for utilization. The phenotypic approach, on the other hand, is necessary for understanding the nature of the breeds of an animal as a consequence of the interaction of their genotype and the environment.

However, the genotypic approach needed for understanding the nature of the animal breeds requires costly laboratory equipment [3]. Moreover, the assertion of using the phenotypic approach in identifying breeds of animals has not yielded satisfactory results due to similarities and variations that exist in animal breeds [3]. Therefore, there is a need for a breed characterization method for precision livestock farming.

This paper seeks to address the above issues by devising an instance segmentation method of Mask R-CNN [4] integrated with Grabcut [5] that can accurately detect and completely extract an object's image during object detection and localization tasks.

Mask R-CNN, an instance segmentation method popular for its class and mask regression has been widely applied for cattle image segmentation tasks [6–10].

2 Motivation

It will be an interesting task to know how to obtain health and welfare information about individual cattle in a typical animal-farm environment. Moreover, knowing the technical know-how interpretation of this information including the activity behaviors of the individual cattle will be more appealing. In addition to the aforementioned information about individual cattle, body condition score and live weight of individual cattle can be obtained including early detection of any illness such as hoof disease and lameness in such an environment using computer vision-based methods. Nevertheless, the available methods are not designed for complete cattle instance segmentation problem in a complex environment. As precise segmentation of each cattle image from its background is a pre-requisite for obtaining the aforementioned information which is highly significant towards obtaining individual cattle information in a real-time environment, there is a need for combined approaches.

3 Related Works

Considerable contributions have been made by different researchers in applying computer vision and image processing algorithms to various image scenes by reiterating the importance of determining specific regions in an image that have intra-class similarity and inter-class diversity for a fine-grained classification [11–15]. In their proposed work, Selvaraju et al. [16] revealed the capability and effectiveness of convolutional neural networks (CNNs) in proposing regions in an image and extracting the features found in such proposed regions, thereby, making the application of CNNs widely accepted for image classification. Other notable works that demonstrated the effectiveness of CNNs in extracting features for the categorization of species are [17–20]. Other tasks in which computer vision is applied are animal identification, detection of animal disease, animal counting, and animal image segmentation [6–9, 21].

Deep learning networks through complex algorithms have tremendously advanced the domain of object detection [22, 23], instance segmentation [4, 24], and semantic segmentation [25]. A bounding box is generated by the majority of object detection methods for each target object detected before classifying these objects [26]. By using selective search, region proposals are generated by the Region-based Convolutional Neural Network (R-CNN) method, and a deep CNN is used for the classification of the object proposals [27, 28]. Nevertheless, it is not economical to extract the features of the proposal region using R-CNN. A separate Region Proposal Network (RPN) is employed in Faster R-CNN [26] for the generation of region proposals. A set of rectangular object proposals is generated as output from the image it takes as input with each having its objectness score [26]. There are two distinct branches in Faster R-CNN: (1) the RPN; this is the first branch which is used in proposing bounding boxes for the candidate object, and (2) the Fast R-CNN detector; this is the second branch which is used in extracting features from each candidate box and performing bounding box regression and classification [27, 28]. Fully Convolutional Networks (FCN) is a CNN variant and segmentation model widely used in image segmentation for transforming pixels of an image to categories of the pixel.

Image segmentation being developed from the object detection seeks to define object classes at a pixel-wise level. Semantic segmentation [25] and instance segmentation [4, 24] are the two categories of image segmentation. The task in semantic segmentation involves pixel-by-pixel labeling of an image for specific object classes; using semantic segmentation, objects from the same class are not distinguished [29].

The approach in instance segmentation involves identifying the instance of each object using a mask representation in the image. Simultaneously, mask extraction and the object class prediction are realized [30]. There are three steps involved in instance segmentation: (1) identification of proposal regions using RPN, (2) prediction of an object class, and (3) extraction of a mask. Through the operation of the convolution, the spatial layout of an input object is encoded by the mask extraction.

Recently, different instance segmentation methods have been proposed for different tasks [31]. Among them is watershed transform, which was combined with modern deep learning to generate an energy map, and, through this, object instance segmentation was realized by a single energy level cutting [32]. Also, a fully convolutional network was extended to actualize the segmentation of beef cattle [6]. A Mask R-CNN instance segmentation method was applied for automatic cattle counting in a quadcopter vision system [8]. Mask-SSD [33] was proposed as a single-stage approach to instance segmentation of objects using a single-shot detector [33].

4 Materials and Methods

The materials and methods used in performing the tasks involved in this study are presented in this section.

4.1 Data Acquisition and Processing

Ten (10) cattle were acquired and employed for the experiments that were performed in this study. Their species were studied to identify their characteristics. 100 images were captured for each cow making the entire image 1000 in number. 200 images were used as a training dataset (20%), and 800 images were used as a testing dataset (80%). To augment the deficiency in the captured dataset and to achieve high accuracy by the deep learning model, the augmentation technique popularly employed in populating pictures was utilized. By applying augmentation techniques such as rotation, shearing, and height and width of image translations, the diversity issue in the acquired dataset was reduced. Furthermore, pre-trained Microsoft common objects in context (MS coco) weights of Mask R-CNN was employed. LabelMe annotation tool [34] was used in annotating the training dataset by mentioning all the coordinates of the object of interest in all the training images for easy prediction of the object's location in the test images by the trained deep learning model.

The dataset of the cattle is a herculean task for the segmentation network to segment when the following aspects are considered:

- The frequent change in position of the cattle object and assumption of different positions especially when moving. To address this, the network is expected to possess a much stronger capacity for generalization.
- The coat patterns resemblance that exists among the cattle with no conspicuous form of identification makes it difficult if not impossible to differentiate between two specific cattle, this, nearly makes the differentiation of one cow from another impossible for human eyes in the event of partial occlusion.

- The impact of illumination variation on the algorithm of machine learning poses a challenge to the entire segmentation process, as machine learning can mistakenly assume patches for cattle's features.
- The impossibility to distinguish between the segmented cow instance and the image's background is also a great challenge in the cow detection task.

Tables 1 and 2 show the model hyperparameters and the model implementation requirements respectively. Figure 1 shows the workflow of the proposed cow image segmentation method.

Table 1 The model hyperparameters

Specifications	Amount in number
Learning rate	0.001
Momentum of learning	0.9
Dimension of image (minimum)	512
Dimension of image (maximum)	512
Detection of confidence (minimum)	0.5
Number of batches	5
Size of batch	200
Epochs	5
Iterations per epoch	5
Mask shape	28×28
Number of anchor classes (cow & background)	2

Table 2 The model implementation requirements

Software	Type/ version	Hardware	Type/version
Operating system	64-bit Windows 10	CPU	Intel Core i5 processor@2.4 GHz
IDE	Visual studio 2019	RAM	16 gigabytes
Python library	Tensorflow [35]	Graphics card	GeForce GTX 1080 Ti
MATLAB	R2019b	Hard-disk	2 terabytes
		Camera module	Vision Datum LEO 640H-200gc High-speed 200fps Sharp RJ33 CCD gigabit ethernet 3d
		Monitor	10.1 inch IPS HD Portable LCD gaming monitor PC display VGA HDMI interface for PS3/PS4/XBOx360/ CCTV/Camera

The loss function employed for this task is according to the following equation:

$$L = L_{ce} + L_{be} + L_{me} \tag{1}$$

where L is the loss function, L_{ce} is the classification error, L_{be} indicates the regression error of the bounding box, and L_{me} is the mask error.

5 Results and Discussion

The obtained segmentation result of the Mask R-CNN instance segmentation method integrated with Grabcut was compared with the result of Mask R-CNN without Grabcut as shown in Fig. 2a–c. The accuracy of the segmentation process is measured as follows:

$$\text{Accuracy (IoU)} = \frac{\text{Area}\,(A \cap B)}{\text{Area}\,(A \cup B)} \times 100 \tag{2}$$

where IoU is defined as the area of intersection by the area of the union of a predicted object bounding box and a ground-truth object bounding box. Table 3 shows the mean average precision (mAP) for the cow image segmentation task.

Figure 3 shows the qualitative result of the Mask R-CNN instance segmentation method [4] integrated with Grabcut [5]. The ground-truth bounding box (in red) was generated using LabelMe annotation tool [34], the predicted cow object bounding box (in green) and the class label (cow) including the prediction

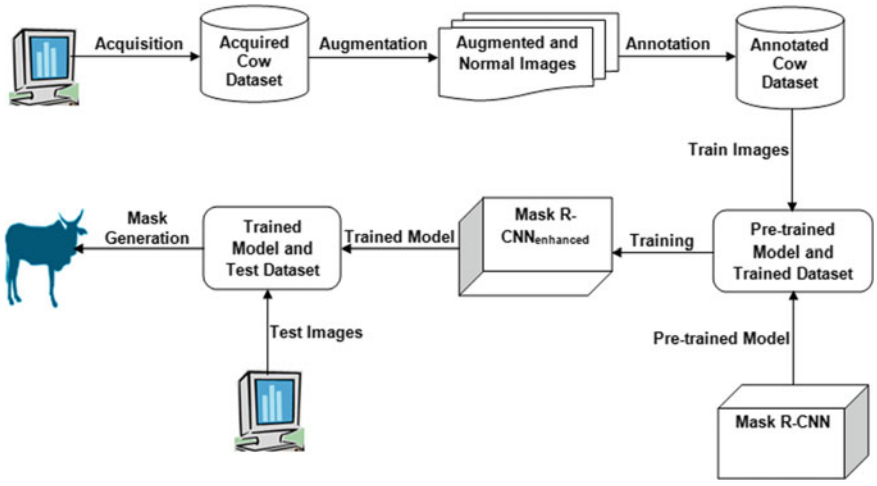


Fig. 1 The workflow of the proposed cow image segmentation method

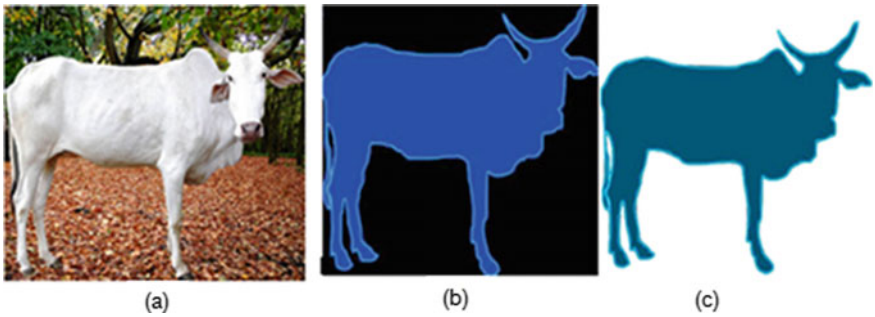


Fig. 2 **a** Cow image with a background. **b** Cow image segmentation using mask R-CNN without grabcut. **c** Cow image segmentation using mask R-CNN integrated with grabcut

Table 3 Comparison of mask R-CNN without grabcut and mask R-CNN integrated with grabcut

Segmentation method	Backbone network	mAP
Mask R-CNN without grabcut	ResNet101	0.92
Mask R-CNN integrated with grabcut	ResNet101	0.93

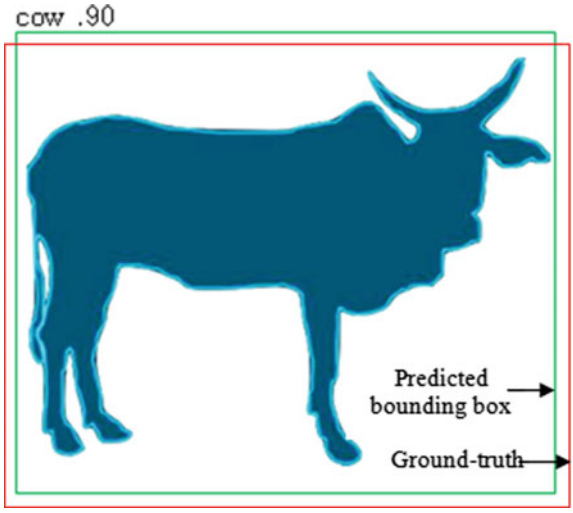


Fig. 3 Result of mask R-CNN instance segmentation method integrated with grabcut

confidence (0.90) were all generated by the bounding box regressor and the object classifier [26–28], and mask [4] was generated for the detected cow object. The integration of Grabcut with the Mask R-CNN instance segmentation method ensures the complete removal of the background binary color in Fig. 2b leading to the complete contour extraction of the cow object as shown in Fig. 2c. In term of accuracy, the proposed model outperforms the accuracy achieved by the recent works [6–8].

6 Conclusion

Segmentation is an important preliminary step in most object detection and classification problems. The choice of an appropriate segmentation method depends on the peculiarity of individual problem characteristics. The problem of identifying a cow object in an image pixel-by-pixel for precision livestock farming was addressed in this paper using Mask R-CNN integrated with Grabcut. The instance segmentation experiment conducted on the acquired cow images produced considerable mAP accuracy with complete extraction of the object from the image without any background, making it fits for the breed's identification task. While the domain of instance segmentation of cow image is well studied, we acknowledge only one previous study that has made use of Mask R-CNN for this similar idea [7]. That study did not utilize Grabcut in their work. Our future work involves localization and identification of multiple and overlapping cow objects in an image using the instance segmentation method.

References

1. FAO (2011) Molecular genetic characterization of animal genetic resources. No. 9 FAO Animal Production and Health Guidelines, Rome
2. FAO (2012) Phenotypic characterization of animal genetic resources. No. 11 FAO Animal Production and Health Guidelines, Rome
3. Ghosh P, Mustafi S, Mandal SN (2020) Image-based goat breed identification and localization using deep learning. Int J Comput Vis Image Process 10(4):74–96
4. He K, Gkioxari G, Dollár P, Girshick R (2017) Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision, pp 2961–2969. IEEE, Venice, Italy
5. Rother C, Kolmogorov V, Blake A (2004) GrabCut: interactive foreground extraction using iterated graph cuts. ACM Trans Graph (TOG) 23(3):309–314
6. Ter-Sarkisov A, Ross R, Kelleher J, Earley B, Keane M (2018) Beef cattle instance segmentation using fully convolutional neural network. arXiv preprint, arXiv:1807.01972, 1–11
7. Qiao Y, Truman M, Sukkarieh S (2019) Cattle segmentation and contour extraction based on Mask R-CNN for precision livestock farming. Comput Electron Agric 165:1–9
8. Xu B, Wang W, Falzon G, Kwan P, Guo L, Chen G, Schneider D (2020) Automated cattle counting using Mask R-CNN in quadcopter vision system. Comput Electron Agric 171:1–12
9. Xudong Z, Xi K, Ningning F, Gang L (2020) Automatic recognition of dairy cow mastitis from thermal images by a deep learning detector. Comput Electron Agric 178:1–11
10. Bello R, Talib A, Mohamed A (2020) Deep learning-based architectures for recognition of cow using cow nose image pattern. Gazi Univ J Sci 33(3):831–844
11. Angelova A, Zhu S (2013) Efficient object detection and segmentation for fine-grained recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 811–818. IEEE
12. Gao S, Tsang IW, Ma Y (2014) Learning category-specific dictionary and shared dictionary for fine-grained image categorization. IEEE Trans Image Process 23(2):623–634
13. Akata Z, Reed S, Walter D, Lee H, Schiele B (2015) Evaluation of output embeddings for fine grained image classification. In: The IEEE conference on computer vision and pattern recognition, pp 2927–2936. IEEE, Boston, MA, USA

14. Xiao T, Xu Y, Yang K, Zhang J, Peng Y, Zhang Z (2015) The application of two-level attention models in deep convolutional neural network for fine-grained image classification. In: The IEEE conference on computer vision and pattern recognition, pp 842–850. IEEE, Boston, MA, USA
15. Xie S, Yang T, Wang X, Lin Y (2015) Hyper-class augmented and regularized deep learning for fine grained image classification. In: The IEEE conference on computer vision and pattern recognition, pp 2645–2654. IEEE, Boston, MA, USA
16. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: The IEEE international conference on computer vision, pp 618–626. IEEE, Venice, Italy
17. Zhang N, Donahue J, Girshick R, Darrell T (2014) Part-based R-CNNs for fine-grained category detection. In: European conference on computer vision, pp 834–849. Springer, Cham
18. Liu X, Xia T, Wang J, Lin Y (2016) Fully convolutional attention networks for fine-grained recognition. arXiv preprint arXiv:1603.06765, 1–10
19. Wang Y, Wang Z (2019) A survey of recent work on fine-grained image classification techniques. J Vis Commun Image Represent 59:210–214
20. Wei XS, Wu J, Cui Q (2019) Deep learning for fine-grained image analysis: a survey. arXiv preprint arXiv:1907.03069, 1–7
21. Bello RW, Talib AZ, Mohamed ASA, Olubummo DA, Otobo FN (2020) Image-based individual cow recognition using body patterns. Int J Adv Comput Sci Appl 11(3):92–98
22. Wang D, Tang J, Zhu W, Li H, Xin J, He D (2018) Dairy goat detection based on faster r-cnn from surveillance video. Comput Electron Agric 154:443–449
23. Norouzzadeh MS, Nguyen A, Kosmala M, Swanson A, Palmer MS, Packer C, Clune J (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci 115:E5716–E5725
24. Pinheiro PO, Collobert R, Dollár P (2015) Learning to segment object candidates. arXiv preprint arXiv:1506.06204, 1990–1998
25. Shelhamer E, Long J, Darrell T (2017) Fully convolutional networks for semantic segmentation. IEEE Trans Pattern Anal Mach Intell 39(4):640–651
26. Ren S, He K, Girshick R, Sun J (2016) Faster R-CNN: towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 39(6):1137–1149
27. Girshick R, Donahue J, Darrell T, Malik J (2015) Region-based convolutional networks for accurate object detection and segmentation. IEEE Trans Pattern Anal Mach Intell 38(1): 142–158
28. Girshick R (2015) Fast R-CNN. In: Proceedings of the IEEE international conference on computer vision, pp 1440–1448. IEEE, Santiago, Chile
29. Li K, Hariharan B, Malik J (2016) Iterative instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3659–3667. IEEE, Las Vegas, NV, USA
30. Zhao K, Kang J, Jung J, Sohn G (2018) Building extraction from satellite images using mask r-cnn with building boundary regularization. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 247–251. IEEE, Salt Lake City, UT
31. Chen L-C, Hermans A, Papandreou G, Schroff F, Wang P, Adam H (2018) Masklab: instance segmentation by refining object detection with semantic and direction features. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4013–4022. IEEE, Salt Lake City, UT
32. Bai M, Urtasun R (2017) Deep watershed transform for instance segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 5221–5229. IEEE, Honolulu, HI
33. Zhang H, Tian Y, Wang K, Zhang W, Wang FY (2019) Mask SSD: an effective single-stage approach to object instance segmentation. IEEE Trans Image Process 29:2078–2093

34. Russell BC, Torralba A, Murphy KP, Freeman WT (2008) LabelMe: a database and web-based tool for image annotation. Int J Comput Vision 77(1–3):157–173
35. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Kudlur M (2016) Tensorflow: a system for large-scale machine learning. In: Proceedings of the 12th {USENIX} symposium on operating systems design and implementation (OSDI '16), Savannah, GA, USA, pp 265–283

Forecasting of Reservoir Inflow Using Machine Learning—Case Study: Klang Gate Dam Reservoir

Karim Sherif Mostafa Hassan, Yuk Feng Huang, Chai Hoon Koo, Tan Kok Weng, Ali Najah Ahmed, and Ahmed Hussein Kamel Ahmed Elshafie

Abstract Climate change is a long-term change in the ordinary weather conditions that affects the local, regional and global climates. One of the response solutions to overcome this phenomenon is by managing water resources more efficiently. The reservoir is a major infrastructure to achieve water resource management and therefore it requires accurate water resources forecasting. The SVR and MLPNN models are introduced as a solution to achieve an efficient reservoir inflow forecasting. There are many input parameters that influence reservoir water flow but the 3 most important parameters are storage level, rainfall, and evaporation that need to be fed into the two models. There have been various model parameters tested such as kernel types in the SVR and the number of hidden layers and neurons in the MLPNN. Both models have proven their ability but however, the MLPNN with two hidden layers and 4 neurons in each layer had outperformed the SVR after being tested using four different performance tests.

Keywords Reservoir inflow · Multilayer perceptrons neural networks · Support vector machine

K. S. M. Hassan (✉) · Y. F. Huang · C. H. Koo · T. K. Weng
Department of Civil Engineering, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Bandar Sungai Long, Cheras, 43000 Kajang, Selangor, Malaysia
e-mail: kareem.hassan@1utar.my

A. N. Ahmed
Institute of Energy Infrastructure, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia

A. H. K. A. Elshafie
Civil Engineering Department, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_4

1 Introduction

1.1 Background of Study

Global warming is the long-term heating of the climate system of the Earth observed during the pre-industrial era (between 1850 and 1900) caused by human activity, mostly due to the combustion of fossil fuels, which raises the amount of heat-retaining greenhouse gases in the atmosphere. A major effect caused by global warming is climate change.

Climate change is a long-term change in the ordinary weather conditions that the local, regional, and global climates of the World have come to describe. This modification has a wide spectrum of effects found that are associated with the impacts of climate change. The resulting impact on water resources, for example has been one that is rather disastrous, worldwide.

First and foremost, it is important to know the critical impacts caused by climate change as referred to the findings found in Intergovernmental Panel on Climate Change's special report published on 2008 [1], and which are highlighted into 3 main points as below:

1. Precipitation has shown a steep increase at high latitudes and at some tropics, and a dramatic decrease at the lower mid-latitude areas.
2. Anticipated that by the mid of 21st-century, water availability will become scarce in dry regions causing severe drought and increases in some wet regions far more than expected causing floods and uncontrollable water flow.
3. Glaciers and icebergs are expected to melt faster causing a rising in water levels and placing some famous world cities in danger of being flooded by water.

All the above impacts have led to freshwater scarcity all around the globe as demonstrated in Fig. 1.

Water resource management may be described as a system for managing resources related to land and water to achieve more social and economic benefits from them. The most critical components of water resource management are

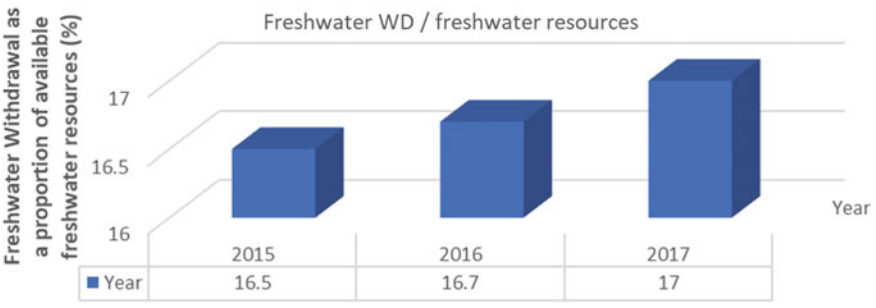


Fig. 1 Global assessment for water scarcity

reservoirs and river basins [2]. Reservoirs are being built to reverse and slow down the effects of global warming and to cater the continuing anthropogenic need for water from a non-continuous natural water supply, and meeting and maintaining human rights to obtain clean water [3]. Despite all these important aspects for reservoir development, the absence of a reliable method for precise measurement of water level is an issue that requires an accurate and precise tool for reservoir forecasting, such as the Artificial Intelligence models.

Reservoir inflow is the most crucial parameter contributing to the amount of water stored. The forecasting and estimation of reservoir inflow is a challenging process as the hydrological parameters recorded and taken into consideration have been increasing. This is due to the highly sensitive sensor systems placed at reservoirs that can detect and store numerous amounts of hydrological parameters that might not have any mathematical linearity among it.

Over the last few decades, physical models, also known as numerical models, have been the conventional method to solve most hydrological problems. Such models require a considerable amount of data about the different parameters representing the dynamic characteristics of hydrological behavior, and these vast data for these parameters must be highly accurate for physical models to work with significant efficiency. Thereafter, the invention of Artificial Intelligence (AI) was then a much-needed solution to these limitations [4].

It is understood that Machine learning algorithms are extremely successful in solving issues that are nonlinear and have no relations among them. AI has thus become more resourceful and preferable in hydrology over traditional physical modeling approaches especially in reservoir parameters forecasting such as inflow, streamflow, etc. [5].

The remainder of the article is arranged as follows. In Sect. 2, the two chosen AI models for the current study which are the Support Vector Regression (SVR) and the artificial neural networks are discussed. In Sect. 3, Klang Gate Dam which is chosen as a case study, and the mathematical equations that build each AI model along with the performance indices to test each model are described in addition to the objectives aimed. In Sect. 4, the reservoir inflow is forecasted, and the results are discussed. Lastly, in Sect. 5 the conclusions are made, and future recommendations for interesting follow up, are given.

2 Literature Review

2.1 Support Vector Regression (SVR)

Support vector regression is classified as supervised learning in machine learning under support vector machine (SVM) that analyze data for regression analysis, and it was firstly introduced in 1997 based on Vladimir Vapnik's concept of support vectors [6]. The SVR operates in a very simple way by maintaining all the key

features defining the optimal margin algorithm hence arranging the data into two main sets; that is the positive sets of data located above the optimal margin and the negative sets of data located below the margins. This categorization process is known as soft margin.

Since the year 2000 support vector regression approach has become popular to be used in a wide range of hydrological problems. In 2011 a novel study has been published testing the SVM model performance by reducing the number of inputs. At first, the SVM is adapted to forecast monthly streamflow by taking into consideration 18 different inputs, the number of inputs is being reduced using three reduction techniques which are the principal component analysis (PCA), Gamma test (GT), and the forward selection (FS) techniques. The PCA and GT have proven their ability in reducing the number of inputs from 18 to just 5 inputs while the FS was able to reduce it from 18 to 7 inputs only. Then these inputs are being fed into the SVM model to test its accuracy. The results were unexpected as the SVM models coupled with the reduction techniques which are the PCA-SVM and the GT-SVM has shown better results compared to the standalone SVM, with 18 inputs as they obtained a score of 0.92 on R^2 ranking skills. It was concluded that these techniques are highly recommended to be fused with the SVM as it improves the performance of the model [7].

The ability of the SVR does not just stop in forecasting reservoir inflow based on hydrology information and ordinary reservoir operation data, but it expands further beyond that. In 2017, a novel study introduced several climate phenomenon indices to be used with the SVR [8]. The SVR was compared to random forest (RF) and Artificial neural network (ANN) models and it was ranked as the second-best model among the three models which proves its strong capability in forecasting reservoir inflow based on various inputs.

2.2 Artificial Neural Network

Initially built to replicate real biological neural systems, particularly the human brain, the Artificial Neural Network (ANN) consists of several interconnected basic computational elements called neurons or nodes, where each node is transmitted with an input signal (this is absolute 'information' from other nodes or external stimuli) and is processed locally through activation or pass-through mechanism. Figure 2 illustrates ANN architecture [9].

To ensure an accurate flow prediction of reservoirs and rivers, correct input with a strong correlation among them should be chosen wisely. A study published in 2010 adapts the Multilayer perceptron (MLP) to forecast daily flow based on historical data and by performing a trial and error method to choose among the best input to perform input parameter sensitivity analysis check on the model. The MLP model is compared to the KNN and the results have shown that the MLP with an architecture of 6-3-1 can forecast the river inflow more accurately when compared to the KNN, obtaining a score value of 0.957 on the R skill score [10].

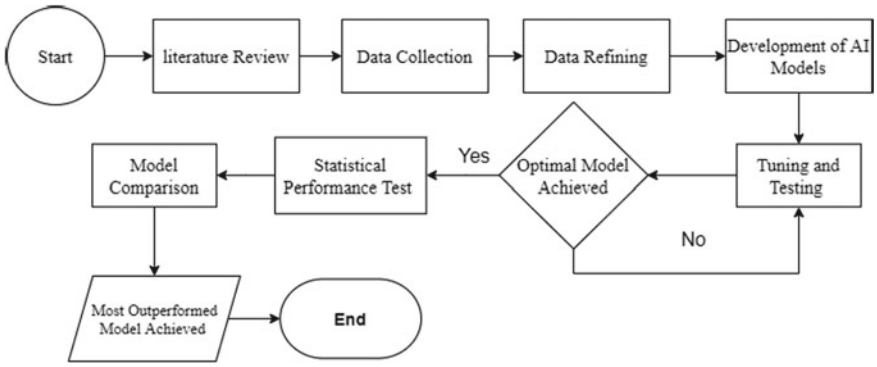


Fig. 2 Research study flowchart

A major issue when forecasting inflow is the inconsistent pattern of water flowing into the reservoir affected by the seasonal rainfall. A study published in 2019 classify the inflow based on high, medium, and low categories and then adapt the MLP to forecast reservoir inflow [11]. Four different models have been developed with different input patterns taking into consideration the seasonal inflow. For fair analysis of the results, the MLPNN has been compared to the SVR under the same input patterns and the results have revealed that the MLPNN is a strong forecasting model but however it is inferior to the SVR.

3 Methodology

3.1 Flow Chart

Literature Review

A comprehensive review on the SVR and MLPNN has been made involving the history for each model and the employment of each model generally in the hydrology field and specifically in inflow forecasting.

Data Collection & Refining

The data has been provided by the Lembaga Urus Air Selangor (LUAS) as KLANG gate dam falls within the Selangor state. The data consisted of four parameters which are rainfall, evaporation, water level, inflow.

Development, Tunning & Testing of AI Model

The two AI models have been developed by feeding them with four inputs and obtaining inflow as the only output. Both of the model's parameters have been calibrated using the trial & error and cross-validation approach. However, in the testing stage the cross-validation has provided poor results and hence the trial-and-error approach was favoured after performance indices were measured.

Model Comparison

The last stage was to compare both model performance in forecasting inflow and to conclude the best performing model.

3.2 Study Area and Data

The data used for this analysis belongs to the Klang Gate Dam (KGD) reservoir and hydrology station in the federal capital of Kuala Lumpur, Malaysia. In 1958 the Klang Gate Dam was known to be the first dam constructed in Malaysia at that time. The KGD properties are found in Table 1 and aerial location are displayed in Fig. 4. The data collected and used for this study consist of 4 main parameters which are rainfall, evaporation, reservoir level, and inflow are for 2000 to 2005. Hence, to further visualize the rainfall trend within Kuala Lumpur Fig. 3 demonstrate the pattern of rainfall. The data is then refined for any errors due to the faulty sensors and the missing data has been updated by replacing it with a new averaged data. The data that is used to feed the AI models are based on two scales, the monthly and daily.

The data was divided into 80% as training sets and 20% for testing the models, the data is split based on the thumb rule as there is no standard way of splitting the data however in ordinary practice the training amount of data should be more than testing data to obtain satisfactory results [12].

Table 1 Klang Gate dam properties

Item	Description
River name	Klang river
Structure type	Concrete gravity dam (Arched)
Date of completion	June 1980
Elevation (EL m)	97.87
Length (m)	138.72
Height (m)	36.89
Location from city centre (km)	12.1
Catchment area (km^2)	77
Flood control volume (m^3) (10^6)	6.2
Flood water level (EL m)	96.92
Total storage capacity (m^3) (10^6)	35.4
Spillway type	Overflow gated
Gate	Slide gate

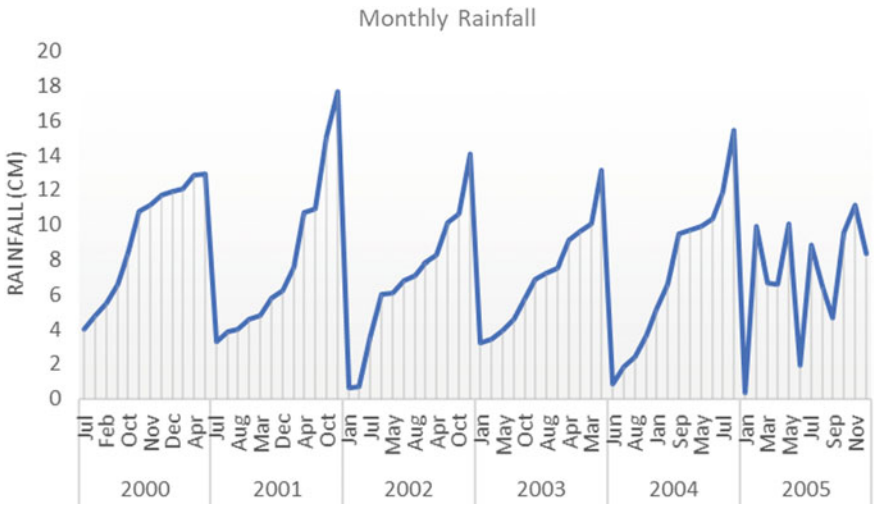


Fig. 3 Monthly rainfall pattern

Fig. 4 Location of Klang Gate dam in Gombak district

3.3 Proposed Inputs for Each Model

$$I = \omega\left[R_f(t)E_v(t)W_l(t)\right] \tag{1}$$

where:

I Forecasted Inflow.
ω Transfer Function.
$R_f(t)$ Rainfall at month (t).
$E_v(t)$ Evaporation at month (t).
$W_l(t)$ Water level at month (t).

Transfer Function for Support Vector Regression

$$\begin{bmatrix} linear{:}\omega(x,xi) = x \cdot xi \\ Polynomial{:}\omega(x,xi) = (Y(x \cdot xi) + r)^d \\ Radial\,Base\,Function(RBF){:}\omega(x,xi) = \exp\left(-\frac{\|x-xi\|}{2\sigma^2}\right) \\ Sigmoid{:}\omega(x,xi) = \tan h(Y(x \cdot xi) + r) \end{bmatrix} \tag{2}$$

Transfer Function for Multilayer Perceptron Neural Network (MLP)

$$\omega = \frac{2}{1 + \mathbf{e}^{\left(\sum_{i=1}^{n_0} \boldsymbol{WijXi} + \boldsymbol{W}_0 j\right)}} - 1\ (\boldsymbol{Hyperbolic\ Function}) \tag{3}$$

$$\omega = \sum_{j=1}^{n_1} vj_k hj + v_{0k}\ (\boldsymbol{Linear\ Function}) \tag{4}$$

3.4 AI Model Statistical Performance Analysis

Mean Absolute Error (MAE)

$$\boldsymbol{MAE} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \left| \frac{\mathbf{y_{pred,i}} - \mathbf{y_i}}{\mathbf{n}} \right| \tag{5}$$

Coefficient (R)

$$\boldsymbol{R} = \frac{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\mathbf{y_o} - \overline{\mathbf{y_o}}\right)\left(\mathbf{y_{f,i}} - \overline{\mathbf{y_{f,i}}}\right)}{\sqrt{\sum_{\mathbf{i=1}}^{\mathbf{n}} \left(\mathbf{y_o} - \overline{\mathbf{y_o}}\right)^2 \left(\mathbf{y_{f,i}} - \overline{\mathbf{y_{f,i}}}\right)}} \tag{6}$$

Mean Square Error (RMSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{pred,i} - y_i)^2 \tag{7}$$

Median Absolute Error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{pred,i} - y_i}{y_i}\right| \tag{8}$$

3.5 Objectives

1. Develop two AI models namely the support vector regression (SVR) and the multilayer perceptron neural network (MLPNN) model to forecast reservoir inflow based on the monthly forecast.
2. Finding the optimum parameters for each model by testing the model with four statistical performance analysis tests
3. Comparison of two models to obtain the better performing model between the two of them.

4 Discussion, Results & Analysis

4.1 Support Vector Regression (SVR)

The first model that has been used to forecast monthly inflow was the SVR. The SVR has a lot of parameters to be considered when running the model through the python programming language. The first parameter is the kernel type and there are four main kernels to be tested in SVR; which are the Radial Basis Function (RBF), Sigmoid, Polynomial and the linear. In addition to the kernel types, there are other parameters that influence the model performance such as the Gamma, the Coef0, and the Epsilon, but however, not all of them affects every kernel. As for the Coef0, it plays a changing factor when used with the sigmoid and polynomial kernels, while it does not have any effect on the RBF and the linear.

The Gamma and Epsilon affect the results when used in all kernels despite having a minor effect on the linear as the linear kernel performance slightly changes when the Gamma and epsilon change. Choosing the parameters is one of the most difficult stages as it has a direct effect on model performance. As for most literature reviews the RBF is said to be the best kernel followed by the sigmoid [13]. Regardless of that, in this study we tested all the four kernels by choosing different

values for Gamma and Epsilon ranging from 0.01 to 0.1 for the Gamma and 0.001 to 0.01 for the epsilon.

The result has been graphically represented in Fig. 5 by plotting the value of R^2 versus Gamma. Despite the model was being tested using four different tests, but we believe that the R^2 skilling score is the clearest among all the four tests and easily understood.

As Fig. 5 shows that the best kernel is the RBF followed by the Polynomial then the linear and lastly the sigmoid. As for the RBF, the Gamma is directly proportional to model performance as when gamma was 0.01 R^2 was 0.53; and when gamma increases R^2 increases as well, until it reaches the optimum values whereby any increase in the gamma after the optimum point will cause a decrease in the performance of the model. This can be seen in the graph; as gamma increase to the optimum point which is 0.05 the R^2 increases to reach the maximum of 0.76 and when gamma increases after 0.06–0.08 the model performance decreases reaching 0.69.

On the other hand, the polynomial acted differently as the gamma increases the R^2 decreases so the maximum point was 0.73 on R^2 when gamma was 0.01; and as gamma increases to 0.1 the R^2 decreases reaching its lowest value of 0.35. The Linear kernel nearly did not show any effect when any of the three parameters changes. It was clearly shown in Fig. 4 as the gamma increases from 0.01 to 0.1 the R^2 stayed almost the same showing a straight line recorded on 0.66. Lastly, the sigmoid kernel has shown an unstable pattern compared to other kernels. As the gamma value increases from 0.01 to 0.05 the R^2 value increases from 0.45 to 0.66 reaching its peak, but as gamma further increases to 0.06 the R^2 drops to 0.57 and when gamma increases again to 0.07 the R^2 shows an increase to reaching to 0.61, but however, any further increase in the gamma beyond 0.07 the R^2 decreases to reach the toe of the graph. Table 2 shows the maximum point achieved and all the parameters used in correlation with that point.

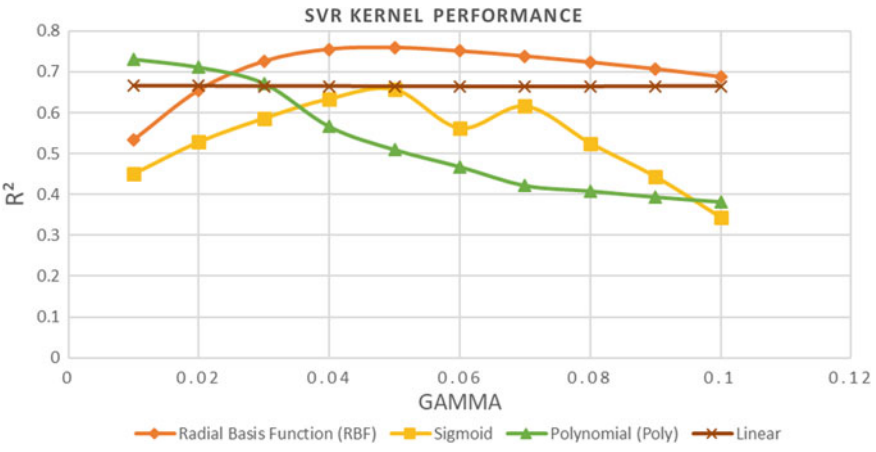


Fig. 5 SVR performance of R^2 versus gamma

Table 2 Best performance of SVR under different parameters

Kernel type	Gamma	Coef0	Epsilon	Mean absolute error	Mean square error	R^2	Median absolute error
Radial basis	0.05	100	0.005000	24.81939776	753.057091	0.75902879	20.36929441
Sigmoid	0.04	100	0.004000	31.66850377	1142.98109796	0.63425676	26.69212034
Polynomial	0.01	15	0.001000	26.10760595	838.5576572	0.731669411	25.52870905
Linear	0.01	100	0.001000	28.52621378	1045.346971	0.665498769	26.78479

4.2 Multilayer Perceptron Neural Network (MLPNN)

The second model that has been used to forecast monthly inflow was the MLP. In the MLP there are a lot of parameters to be trained and chosen and hundreds of combinations to be tried. First and foremost, to ensure a proper running of the MLP model in python the minimum iteration should be at least 500 iterations and above if your data is relatively large to avoid error runout and incomplete conversion of the model. Secondly, the most important parameter is the activation function as it has a major influence on the model since it is in charge to activate the function between the hidden layer and the output layer. There are four types of activation functions; which are firstly the identity function, and it is the most recommended one. This is then followed bt the logistic function and then followed by the relu function and lastly by the tanh function.

The third parameter is the solver which is directly associated with the activation function when seeking an optimum model performance. There are three types of solver which are the lbfgs, sgd, adam. Furthermore, the tol and alpha influence the Bias of the model which affects the model performance, despite all these parameters mentioned above, the most crucial variable that has the major effect on MLP is the number of hidden layer and node sizes. There is no standard way in choosing the number of hidden layers and neurons. However, the one adopted in this study is based on [14] by defining the number of inputs and then adding one to it to be the number of hidden layers. Therefore, since in this study there are 3 inputs so the number of the hidden layer are defined as 4 and the number of neurons is 4. In this study, the best activation function will be demonstrated which is the Identity; and Fig. 5 will present the different solver functions along with tols effect on model performance.

The results presented in Fig. 6 are for 3 different solve functions tested with a range of Tol of 0.0001, 0.001, 0.01, 0.1, 1. The three types of solver functions share the same trend in which as the Tol increases from 0.0001 to 0.1 the model performance measured by R^2 decreases, however, the declination in the R^2 is not of the same decrement. The peak point was recorded by the LBFGS solver in which it had a Tol of 0. 1 and an R^2 of 0.87 despite the hidden layer for this trial was not (4) but it was (4,4) which shows that there were two hidden layers used with 4 neurons in each layer. Regardless of that the pattern recorded by one hidden layer of 4 neurons (4) shown that when Tol increases the value model performance decreases within a positive range. On the other hand, the sgd had an optimum point of 0.83 while the adam solver had an optimum point of 0.77 recorded with a Tol of 0.0001 and any further increase in the Tol causes the model performance to drop dramatically to record an R^2 of below 0 which shows overfitting of the model. Table 3 shows all the optimum values and parameters recorded for the Identity activation function.

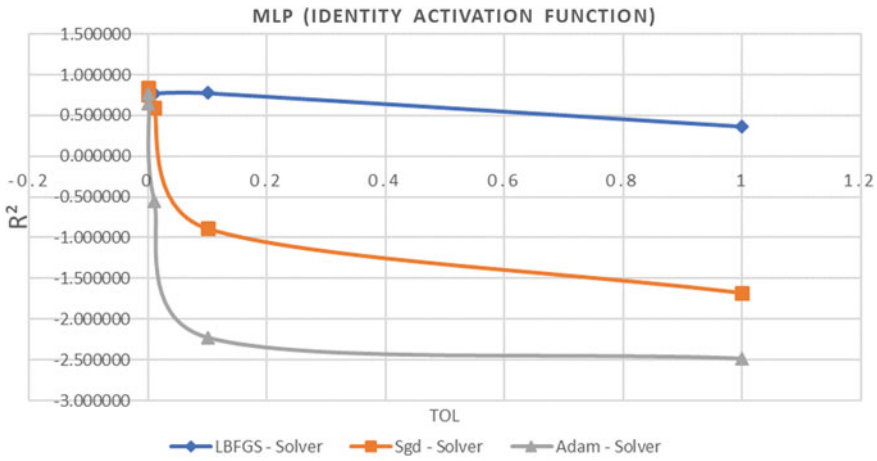


Fig. 6 MLP performance of R^2 versus Tol

5 Conclusion

The present study aimed to develop two artificial intelligence model which can forecast reservoir inflow with a minimum number of inputs. Two models namely the SVR and the MLPNN were developed successfully and optimum parameters were found despite that these parameters are just applicable for this case study and it varies from one case study to another.

When the two models are compared with each other, the MLPNN had obtained a better overall performance score than the SVR; when tested by mean average error (MAE), mean squared error (MSE), coefficient of correlation R^2, median absolute error (MDAE). In a nutshell, the MLPNN has outperformed SVR. However, as a future recommendation, both models should be fused with optimization techniques to avoid the trial-and-error method when defining the parameters and to improve the model performance.

As a future recommendation, further study could be made by developing hybrid models using optimization algorithms such as genetic algorithm (GA) or particle swarm optimization algorithm (PSO) etc. Besides, more inputs could be added to improve the accuracy of the model such as involving seepage.

Table 3 Best performance of MLPNN under different parameters

Max_iterations	Activation	Solver	Hidden layer and nodes size	Tol	Alpha	MAE	MSE	R^2 eror	MDAE
1000	Identity	Ibfgs	4	0.1	1	21.367443	701.498824	0.775527	14.079119
			(4,4)	0.1	1	18.612976	396.757904	0.873041	15.647414
		sgd	4	0.0001	1	20.008538	504.518939	0.838559	20.222036
		adam	4	0.0001	1	25.097602	737.545306	0.763992	28.184304

References

1. Bryson B, Zbigniew WK, Jean P (2008) Technical paper of the intergovernmental panel on climate change. IPCC Secretariat, Geneva, Intergovernmental Panel on Climate Change, pp 210. ISBN: 978-92-9169-123-4
2. Liu J, Yang H, Gosling SN et al (2017) Water scarcity assessments in the past, present, and future. Earth's Futur 5:545–559. https://doi.org/10.1002/2016EF000518
3. Valizadeh N, El-Shafie A, Mirzaei M et al (2014) Accuracy enhancement for forecasting water levels of reservoirs and river streams using a multiple-input-pattern fuzzification approach. Sci World J. https://doi.org/10.1155/2014/432976
4. Dibie Chidubem Damian (2019) A critical review on artificial intelligence models in hydrological forecasting how reliable are artificial intelligence models. Int J Eng Res V8. https://doi.org/10.17577/IJERTV8IS070123
5. Choong SM, El-Shafie A (2014) State-of-the-art for modelling reservoir inflows and management optimization. Water Resour Manag 29:1267–1282. https://doi.org/10.1007/s11269-014-0872-z
6. Drucker H, Surges CJC, Kaufman L et al (1997) Support vector regression machines. Adv Neural Inf Process Syst 1:155–161
7. Noori R, Karbassi AR, Moghaddamnia A et al (2011) Assessment of input variables determination on the SVM model performance using PCA, gamma test, and forward selection techniques for monthly stream flow prediction. J Hydrol 401:177–189. https://doi.org/10.1016/j.jhydrol.2011.02.021
8. Yang T, Asanjan AA, Welles E et al (2017) Developing reservoir monthly inflow forecasts using artificial intelligence and climate phenomenon information. Water Resour Res 53:2786–2812. https://doi.org/10.1002/2017WR020482
9. Reilly DL, Cooper LN (1995) An overview of neural networks: early models to real world systems. 300–321. https://doi.org/10.1142/9789812795885_0023
10. Eskandarin A, Nazarpour H, Teimouri M, Ahmadi MZ (2010) Comparison of neural network and K-nearest neighbor methods in daily flow forecasting. J Appl Sci 10:1006–1010
11. Allawi MF, Jaafar O, Mohamad Hamzah F, El-Shafie A (2019) Novel reservoir system simulation procedure for gap minimization between water supply and demand. J Clean Prod 206:928–943. https://doi.org/10.1016/j.jclepro.2018.09.237
12. Zhang X, Yu X, Qin H (2016) Optimal operation of multi-reservoir hydropower systems using enhanced comprehensive learning particle swarm optimization. J Hydro-Environment Res 10:50–63. https://doi.org/10.1016/j.jher.2015.06.003
13. Zhang D, Lin J, Wang D et al (2018) Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. J Hydrol 565:720–736. https://doi.org/10.1016/j.jhydrol.2018.08.050
14. May DB, Sivakumar M (2009) Prediction of urban stormwater quality using artificial neural networks. Environ Model Softw 24:296–302. https://doi.org/10.1016/j.envsoft.2008.07.004

An Unsupervised Semantic Model for Arabic/French Terminology Extraction

Setha Imene and Aliane Hassina

Abstract We provide in this work, a new approach for automatic bilingual terminology extraction involving Arabic and French languages. We aim to build a bilingual list of terms for a specific field, namely "Algerian Culture". The approach is based on building a semantic space for context understanding, adopting COALS (Correlated Occurrence Analogue to Lexical Semantics). COALS is an interesting semantic model which combines the advantages of LSA and HAL models. We use "Wikipedia" to build a comparable corpus for "Algerian heritage and culture" and a translation tool for term alignment.

Keywords Bilingual terminology extraction · COALS · Machine learning · Comparable/parallel corpus · Natural language processing

1 Introduction

Terms are the main building blocks of a text, their combination gives the whole context and presents its different ideas. Terms play a crucial role for different NLP tasks in every possible field (Bioinformatics, medicine, technology…etc.) such as information retrieval, thesaurus construction, text classification, text indexing, text summarization, and more particularly machine translation which faces an increasing demand for even lower costs and efforts. Approaches proposed in the literature for terminology extraction are linguistic, statistical, machine learning or hybrid approaches. In this work, we propose an unsupervised semantic based approach which uses the COALS model. COALS [1] combines the advantages of two widely used semantic models: LSA and HAL and hence, captures context of both words and documents. We work on an Arabic/French comparable corpus which is extracted from Wikipedia in the specific field of "Algerian heritage and culture". The rest of the paper is organized as follows: Sect. 2 presents related works in bilingual terminology extraction. Section 3 describes our methodology using the

S. Imene (✉) · A. Hassina
CERIST, Research Center on Scientific and Technical Information, Algiers, Algeria

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_5

COALS semantic space model. In Sect. 4, we describe our alignment method. Experiments and evaluation are presented in Sect. 5. Following a discussion, a conclusion ends this paper.

2 Background and Related Works

Terms can be extracted from parallel or comparable corpora. Approaches proposed in the literature can be classified as linguistic, statistical, machine learning, graph-based and hybrid methods.

Traditionally, **linguistic approaches** utilize term's properties such as syntactic and morphological patterns to extract terms such as POS-tagging, chunking…etc. To cite some relevant works, [2] built their own general POS-tagger to extract terms in 21 languages, their main idea was to create a term grammar for each language using the "EuroVoc"parallel corpus. [3] used chunking to build a glossary of terms using Noun-Phrase Chunking. Another technique is the initial seed dictionary introduced by [4] who constructed their own dictionary for English-German languages to extract terms according to their spelling features.

Statistical approaches are language independent and based on statistical measures by measuring the degree of specificity (termhood and unithood).These methods were first proposed by [5] who used a language independent statistical method without any syntactic or semantic model. Statistical methods rely on: term frequencies, using TF-IDF measures for ranking [6], term co-occurrence context like in [7] who used the "NC-value". Moreover, we find these measures in Latent Semantic Analysis LSA [8], in topic modeling [9] and LDA in [10] for extracting terms according to their contextual distributions.

Recently most extractors are **Machine learning** or **Deep Learning** Based, using huge training data. For instance, using an SVM binary classifier, [11] performed term extraction in over 20 languages on EUROVOC thesaurus. Logistic Regression model LIBLINEAR in [12] was proposed to classify if a word is actually a term or not, investigating different types of features. Reference [13] proposed a method called end-to-end nested term extraction using both classification and ranking to distinguish whether terms are conceptual for a specific domain. They use both CNN and LSTM neural networks to improve their results. But the main problem facing these approaches is the availability of data.

Graph based methods are used in ABTE, mainly to propagate seed dictionaries. Reference [14] presented a novel graph-based method called label propagation where each word is represented by a distribution of translated seeds. The seed distributions are propagated over a graph representing relations among words. The authors used two types of graphs: a co-occurrence graph, representing co-occurrence relations between words and a similarity graph, representing context similarities between words. Finally, translation pairs were extracted by identifying word pairs with a high similarity in the seed distributions.

Combining all previous works and methods, the **hybrid approach** for term extraction [15] is a combination of two or more of the above approaches. Reference [16] proposed a simple and effective approach for extracting a bilingual lexicon from a comparable corpora enhanced with a parallel corpora. They propose a new technique to study how parallel sentences extracted from a comparable corpus can improve the quality of candidate translations.

For **Arabic language**, many interesting works were proposed. Reference [17] proposed a hybrid method for the acquisition of a list of domain-specific terms from the Arabic text of the Quran using Abbas dataset.[1] Reference [18] applied linguistic and statistical filtering. Linguistic filtering using a POS-tagger and a sequence identifier, a statistical filtering step incorporates contextual information with a new proposed association measure based on Termhood and Unithood estimation named NTC-Value.

3 Proposed Model and Methodology

From the state of the art review on bilingual term extraction, we notice that very few works rely on extracting terms based on their semantic relatedness. We propose in this work an unsupervised approach based on COALS semantic model which will be detailed in the following subsections. The general approach is depicted in Fig. 1.

3.1 Monolingual Term Extraction

Building a term extraction system consists of two main steps, monolingual extraction and bilingual alignment. We first introduce monolingual extraction, which is conducted through three main steps:

term Identification: text analysis, preparing a list of candidate terms for the extraction.

candidate selection: filtering candidate terms.

candidate ranking: using statistical measures such as Tf-Idf, LLR, C-value… etc.

Our method requires a bilingual corpus for both source and target languages.

Corpus, Data Collection and Preparation: A good term extraction system is directly related to the quality of the chosen corpus and for Arabic language, few parallel corpora are available. Therefore, we use the largest freely available resource, that is the online "wikipedia" as our comparable corpus. In addition to its huge number of articles in different fields, wikipedia is known for its different types of hyperlinks that may contain various information: inter-language links, back/

[1] Abbas dataset: Classical Arabic corpora for religious education and understanding.

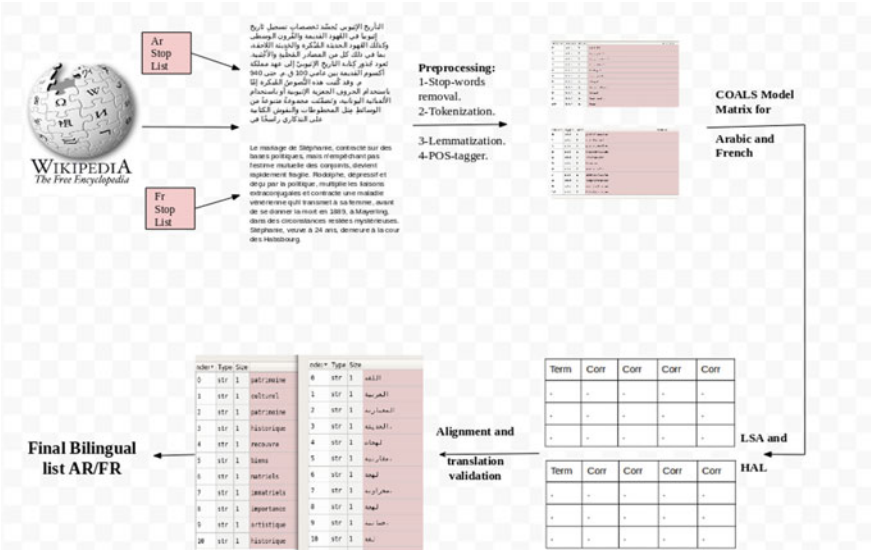


Fig. 1 System overview

forward links, text links. Making use of this wikipedian architecture and supposing that each author relates original articles to their right place in other languages [18], we initiate the process for both languages (Ar/Fr) with one title that extends for all subtitles and we build our corpus for the "Algerian heritage and culture" field stopping when we have enough size of data. The wikipedia corpus is available as XML and SQL files.[2]

Preprocessing and Normalization: While parsing data, a normalization step is performed to clean all texts and leave only important information using tokenization, stop-words removal, lemmatization, diacritics and special characters removal and text annotation (POS-tagging). We used the nltk toolkit[3] for previous normalization and Stanford POS-Tagger[4] for text annotation. At the end of the preprocessing step we obtain two lists of terms for both Arabic and French languages.

Statistical Filtering: Statistical measures are used to select terms after preprocessing. They're based on term frequency, correlation and regression…etc. Statistical measures have proven their strength in different NLP tasks where several techniques are used such as Tf-Idf, C-value, Log-Likelihood…etc. We use TF-IDF measure in order to compute the importance of a term in the text by computing the occurrence frequency in the corpus and the document. We obtain a weighted term

[2] Wikipedia dumps: https://lists.wikimedia.org/pipermail/xmldatadumps-l/.

[3] NLTK toolkit: https://www.nltk.org/.

[4] Stanford POS-Tag: https://nlp.stanford.edu/software/tagger.html.

list ranked in descending order from the most frequent term to the least frequent one. These TF-IDF weights are used in The COALS model Matrix which we describe below.

3.2 COALS Model

COALS belongs to the family of semantic models, also called psychological models. Semantic models are based on the concept of semantic space and are used for many complex unsupervised machine understanding tasks (word recognition, priming, and higher-order comprehension processes). Briefly explained, a semantic space is a matrix whose rows correspond to target words and columns to dimensions that are considered as basic elements. Distance similarity between words is counted using any statistical distance measure [19]. Semantic models investigated in this work are based upon the idea that the word's meaning is related to the context in which the word is usually used according to the assumption that "Semantic similarity determines the relation between terms, where two (lexically) different terms share a similar meaning if they occur in similar contexts" [19]. These models take as context a small number of unordered words, representing each of them as a single vector. We obtain a segmented corpus presented as a set of vectors of different sizes. These vectors are constructed in such a way that words similar in meaning will have closely similar vectors. Semantic similarity may be computed using any measure. Our approach is based on the COALS model to improve Bilingual Term Extraction. COALS has been proposed by [1] and we have chosen it because it combines the advantages of two widely used semantic models: Hyperspace analogue to language (HAL) [20] and Latent semantic analysis (LSA) [1]. According to [1] COALS model predicts human semantic similarity judgments better than HAL, and in most cases better than LSA or other models and scored results confirm that COALS provides a better fit to behavioral word-pair similarity ratings and multiple choice vocabulary tests [1]. Our COALS algorithm is summarized in the following steps:

1. Build the initial matrix and fix ramp-window K = 4.
2. Normalize the matrix and reduce columns.
3. Compute correlation using Pearson's measure.[5]
4. Discard negative values, and root square positive values.
5. Reduce the matrix using SVD function.
6. Compute cosine similarity between word vectors.

Since LSA is a document-based semantic space model and HAL is a word-based semantic space model, COALS captures the context for both words and documents.

[5] Pearson's correlation measure (Karl Pearson 1880).

3.3 Terms Translation (Alignment)

The second part of this work is dedicated to the alignment process. The main assumption for the alignment task is that: a term and its translation share the same contextual similarity. In other words, it is the process of mapping two words in two sentences in two different languages. Most alignment methods use parallel corpora for simple/multiple term translation, we can classify them into the following:

Dictionary-based alignment: These methods use a bilingual dictionary to translate target and source documents and a comparison process is then performed using similarity measures.

Statistical-based alignment: Also called probabilistic methods are the most known techniques for alignment. They use IBM statistical alignment models that are known for their quite good performance in matching tasks.

Machine learning-based alignment: The Alignment process is treated as a classification problem. A classifier is used to find the translation pair from source to target by maximizing the conditional probability between terms.

In deep learning, the general approach using neural networks is based on word embeddings to understand the semantics of terms vectors, the alignment is performed by comparing the sequences of these representations and matching them.

Several alignment tools are proposed in the literature: Giza++, Pialign, Transbank alinger, LF aligner, HunAlign and many others. These tools use different criteria like term frequencies, positions and length in order to find the best equivalent of a term.

In this work, we use semantic information from a comparable corpus and we aim to find the best aligner to translate our comparable terms between French and Arabic languages. We choose the Google translate API to automatically translate our list of Arabic extracted terms into French, then validate those translations. **Google translate API**[6] can be freely used for 78 languages. It is known that Google-translate API uses semantically phrase-based alignment resulting in top translation, definition, and alternative equivalents of a given term.

4 Experimentation and Results

For the bilingual term extraction task, there are no absolute standards for evaluation due to the number of methods and techniques one can choose. Generally, a manual evaluation is performed on a number of extracted terms. Hence, in order to simplify this task of evaluation, we divide it into two parts: firstly, we evaluate the quality of extracted terms from the COALS model using the precision metric. Secondly, we evaluate translation quality by comparing it to other tools and methods.

[6] https://pypi.org/project/googletrans/.

4.1 Evaluation Dataset

From the heritage (“Patrimoine”, “ التراث”), specialized domain, we choose five Wikipedia articles for the evaluation task for both Arabic and French languages:

“تراث ثقافي”, “التراث المغربي”, “الإرث الوطني”, “قيم وموروثات”, “حضارة”

“Civilisation”, “Patrimoine culturel”, “Arts”, “Traditions”, “Héritage national”. We create a small data sample as shown in Table 1.

From the Wikipedian data, we notice that the number of tokens in French is much more than Arabic. Therefore, we split the corpus and select the Top 100 terms from our COALS model results for each language to use in evaluation. We note that the initial specialized Wikipedia corpus provides us to work on two lists in both languages with a number of 530,743 token for Arabic and 290,655 for French.

4.2 COALS Extraction Results

The process of extracting terms Using COALS model produces a huge matrix with normalized correlation degrees between each and every term (see Fig. 2). Using python libraries such as numpy and pandas, we build an organization of relevant accessible terms inside the dataset. Terms are presented in the following figure including related information. Terms that are relevant are stored in the correlation matrix, others are automatically ignored and deleted.

For instance, the first range of both Arabic and French corpus presents 10 relevant and in-domain terms:

For Arabic: “ “مغاربية”, “حديثة”, “لهجات”, “للبلاد”, “اللغة”, “عربية”, “الجزائر””.

For French: Culturel—Patrimoine—Historique—Biens—Matériels—Importance—Artistique. To examine our system performance, we apply the precision measure on this list of extracted terms using COALS model results. We calculate the rate of well recognised terms as a fraction of the number of true valid terms on the number of all the terms. Table 2 shows the Top 100 terms’ precision results with and without preprocessing for French and Arabic languages. We record the same rate for both languages, using the same tool (NLTK) for top 100 terms of our corpus.

Table 1 Number of tokens

Language	Number of tokens
Arabic	2014
French	12,746

Index	culturel	patrimoine	historique	biens	matriels	im
culturel	1	0.400544	0.122936	0.0703838	0.214922	-0.0
patrimoine	0.400544	1	0.0594714	0.0463093	0.138139	-0.0
historique	0.122936	0.0594714	1	0.0343012	-0.0421824	0.29
biens	0.0703838	0.0463093	0.0343012	1	0.00572727	0.04
matriels	0.214922	0.138139	-0.0421824	0.00572727	1	0.08
importance	-0.0377421	-0.0190113	0.298596	0.042738	0.0825383	1

Fig. 2 Correlations between terms

Table 2 Precision results

Extracted terms	Precision for Arabic	Precision for French
Without preprocessing	70%	70%
With preprocessing	95%	95%

4.3 Translation Validation

After examining COALS model performance for the task of extraction, we have to evaluate resulting translations of the system. In Table 3, we show the top 5 terms translations using different translation tools, in comparison with Google Translate API. We use Three available tools to compare their performance against each other:

Google Translate API: A powerful translation platform from Google, that supports around 104 languages.

Yandex: Russian self-learning statistical machine translation system for term based translation supports 90 languages.

Sogou: a very powerful Chinese translation tool which uses Deep earning algorithms. Started basically for (Chinese/English) pair of languages, and now supporting more than 60 languages around the world.

Table 3 Translation tools precision results

Google-translate API	Sogou-API	Yandex-API
Langue (s)	Langue (langues)	Langue (s)
Catégorie	Classification	Cote
Algérie	Algérie	L'algérie
la langue arabe standard	Arabe standard	Arabe architectural
Dialectes modernes du Maghreb	Dialectes maghrébins modernes	Accents modernes maure

- We use Arabic as source language and French as target language to translate the following list of terms:

“

“اللغة (اللغات)”, “تصنيف”, “الجزائر”, “اللغة العربية المعيارية”, “لهجات حديثة مغاربية”

”

We manually check and validate results from each of the three tools according to their context.

5 Discussion and Observations

The first step of evaluation shows the great impact of preprocessing and statistical filtering on the used dataset as shown in Table 2. According to the good and satisfying results we get using the semantic model COALS, we proved that semantic information of terms can be very useful for extracting more relevant and correct terms especially for Arabic. COALS model clearly outperforms other methods in Arabic language term extraction.

The second experiment describes different tools performance for terms translation, obviously Google-translate outperforms other tools with 85% of precision on translated terms. we notice that google translate gives higher results for multiple word terms, unlike Yandex which is a word based translator. On the other hand, Sogou gives amazing results almost 100% true, still the main disadvantage is that it takes much more time to execute one operation compared to the others. This is the first work which uses the COALS model in ABTE and the results confirmed our intuition in choosing it for the task. Indeed, it outperformed state of the art approaches for this task, as shown in Table 4.

6 Conclusion and Future Works

We presented in this paper a new approach to extract bilingual Arabic/French terminology from a comparable corpus. We chose the COALS model for extracting relevant terms in the desired languages because it combines advantages of LSA and HAL. COALS model was applied on a “cultural” comparable corpus automatically built from Wikipedia. It showed a great performance in terms of relevancy of extracted terms. Experiments show clearly that applying multiple filters on the corpus raises precision values in a remarkable way. Furthermore, by comparing several alignment tools, we get to use the best translation tool for validation. As future work, we plan to experiment neural models and compare the results regarding traditional semantic models such as COALS and also compare or develop new alignment tools.

Table 4 Brief comparison of ABTE methods

/	Method	Source/target languages	Domain of application	Precision
[21]	Hybrid	Arabic/Italian	International law	70–80%
[22]	Hybrid	Arabic/English	Newspaper archives	80–90%
[23]	Statistical	Arabic/English	Qur'anic text	74%
–	Hybrid: probabilistic	Arabic/English	Islamic corpus	70–75%
COALS	Hybrid	Arabic/French	Wikipedia	95%

References

1. Rohde DLT, Gonnerman LM, Plaut DC (2006) An improved model of semantic similarity based on lexical co-occurrence. Commun ACM 8:627–633
2. Aker A, Paramita ML, Barker E, Robert D, Gaizauskas J (2014) Bootstrapping term extractors for multiple languages. In: LREC proceedings, pp 483–489
3. Arora C, Sabetzadeh M, Briand L, Zimmer F (2014) Improving requirements glossary construction via clustering: approach and industrial case studies. In: Proceedings of the 8th ACM/IEEE international symposium on empirical software engineering and measurement, pp. 1–10
4. Koehn P, Knight K (2002) Learning a translation lexicon from monolingual corpora. In: Proceedings of the 2nd ACL workshop on unsupervised lexical acquisition, pp 9–16
5. Fung P (1998) A statistical view on bilingual lexicon extraction: from parallel corpora to non-parallel corpora. In: Conference of the association for machine translation in the Americas. Springer, pp 1–17
6. Eck M, Stephan V, Waibel A (2005) Low cost portability for statistical machine translation based on n-gram frequency and tf-idf. In: Proceedings of IWSLT
7. Frantzi K, Ananiadou S, Mima H (2000) Automatic recognition of multi-word terms: the c-value/nc-value method. Int J Dig Libr 3(2):115–130
8. Gaussier E, Renders JM, Matveeva I, Goutte C, Déjean H (2004) A geometric view on bilingual lexicon extraction from comparable corpora. In: Proceedings of the 42nd annual meeting of the association for computational linguistics, pp 526–533
9. Vulic I, De Smet W, Moens MF (2011) Identifying word translations from comparable corpora using latent topic models. In: Proceedings of the 49th annual meeting of the association for computational linguistics, pp 479–484
10. Li S, Li J, Song T, Li W, Chang B (2013) A novel topic model for automatic term extraction. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval, pp 885–888
11. Aker A, Paramita ML, Gaizauskas R (2013) Extracting bilingual terminologies from comparable corpora. In: Proceedings of the 51st annual meeting of the association for computational linguistics, pp 402–411
12. Judea A, Schütze H, Brügmann S (2014) Unsupervised training set generation for automatic acquisition of technical terminology in patents. In: Proceedings of the 25th international conference on computational linguistics, pp 290–300
13. Gao Y, Yuan Y (2019) Feature-less end-to-end nested term extraction. In: CCF international conference on natural language processing and Chinese computing, pp 607–616
14. Tamura A, Watanabe T, Sumita E (2012) Bilingual lexicon extraction from comparable corpora using label propagation. In: Proceedings of the joint conference on empirical methods in natural language processing and computational natural language learning, pp 24–36

15. Ventura JA, Jonquet C, Roche M, Teisseire M (2014) Biomedical terminology extraction: a new combination of statistical and web mining approaches. In: Proceedings of JADT: Journées d'Analyse statistique des Données Textuelles, pp 421–432
16. Morin E, Prochasson E (2011) Bilingual lexicon extraction from comparable corpora enhanced with parallel corpora
17. Alrehaili S, Atwell E (2016) A hybrid-based term extraction method using the Arabic text of the qur'an. In: 4th international conference on islamic applications in computer science and technologies (IMAN 2016), pp 20–22
18. Boulaknadel S, Daille B, Aboutajdine D (2008) A multi-word term extraction program for Arabic language. In: LREC. Citeseer
19. Sellami R, Sadat F, Hadrich Belguith L (2012) Extraction de lexiques bilingues a partir de wikipédia. In: JEP-TALN- RECITAL 2012, workshop TALAf 2012: Traitement Automatique des Langues Africaines (TALAf 2012: African language processing), pp 107–117
20. Burgess C, Livesay K, Lund K (1998) Explorations in context space: words, sentences, discourse. Discourse Process 25(2–3):211–257
21. Fawi F, Delmonte R (2015) Italian-arabic domain terminology extraction from parallel corpora. In: CLiC it, p 130
22. Sabtan Y (2016) Bilingual lexicon extraction from Arabic-English parallel corpora with a view to machine translation. Arab World Engl J (AWEJ), special issue on translation (5)
23. Lahbib W, Bounhas I, Slimani Y (2018) A possibilistic approach for Arabic domain terminology extraction and translation. In: International symposium on computer and information sciences. Springer, pp 231–238

Research on Improving E-commerce Logistics Service Customer Satisfaction Through Application of AGV in Intelligent Warehouse

Ying Xi, Noorsiah Binti Ahmad, and Abdullah Al Mamun

Abstract In the process of e-commerce logistics operation, warehousing order picking speed directly affects customer satisfaction. With the surge of orders, e-commerce intelligent warehouses apply automatic guided vehicles (AGV) to improve the efficiency of order processing and increase customer satisfaction. This paper proposes using simulation to optimize the AGV road network through building an intelligent warehouse simulation model in Anylogic simulation software to collect data of AGV operation under different road networks, and optimize the road network by changing the warehouse layout. Moreover, in order to cope with the rapid increase in the number of orders, this paper builds an AGV order batching integer planning model. Through applying the optimized intelligent warehousing AGV road network, combined with the integer programming model and algorithm for order batching, e-commerce intelligent warehouse can effectively increase the efficiency of AGV order picking, reduce AGV operating costs, therefore improve e-commerce logistics service quality and customer satisfaction.

Keywords AGV · Intelligent warehouse · Road network optimization · Order batching · E-commerce customer satisfaction

1 Introduction

In the past 10 years, e-commerce has a massive growth, especially in 2020. Due to the lockdown caused by the COVID-19 pandemic, the number of orders processed by e-commerce has surged. Based on movement control order (MCO) and online

Y. Xi (✉) · N. B. Ahmad · A. Al Mamun
Faculty of Business and Management, UCSI University, 56000 Kuala Lumpur, Malaysia
e-mail: 1001955415@ucsiuniversity.edu.my

N. B. Ahmad
e-mail: noorsiahahmad@ucsiuniversity.edu.my

A. Al Mamun
e-mail: abdullaham@ucsiuniversity.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_6

shopping platform promotions, in the first quarter of 2020, Malaysia's parcel delivery volume increased significantly to 61 million, an increase of 9% from 56 million in the first quarter of 2019. The parcel volume is even closer to 100 million in the second quarter of 2020 [1]. In China, the 2020 double eleven (11.11) e-commerce shopping festival held by Taobao and Tmall, which are China's most popular e-commerce platforms of Alibaba Group, has a total of 2.321 billion orders in 11 days. This is the 12th double eleven e-commerce shopping festival, and the orders started from 260,000 since 2009, with an increase of nearly 9,000 times in 12 years [2].

When the e-commerce logistics distribution center processes orders, the picking operation consumes major labor and occupies most of the distribution center operation time. According to Bogue [3], e-commerce fulfillment is essentially a piece-picking operation and on-line retailers are having to change and update their warehousing and delivery operations to meet growing demand and customer expectations for a fast and efficient service. The efficiency of the picking operation will directly affect the operation efficiency of the distribution center, and thus affect the customer service level and customer satisfaction. The traditional distribution center adopts the human-to-goods picking mode in which pickers shuttle back and forth in the shelves, which is labor intensive, time-consuming, and easy to make mistakes in the picking process. In order to better meet the development needs of e-commerce companies, it is essential to improve the efficiency of picking operations in distribution centers. Therefore, shortening the order picking time is an effective way to improve logistics efficiency, reduce logistics costs, and improve customer satisfaction.

Logistics management has been seeking mechanized and intelligent handling technology and equipment to reduce handling costs and improve material handling efficiency. Among these automation technologies and equipment, automatic guided vehicle (AGV) is the most flexible system that can meet the needs of complex material handling. The intelligent warehouses based on AGV have emerged in recent years. Pickers stand in front of a fixed workbench, and storage robots move the shelves to the pickers to complete order picking. This new type of goods-to-human picking operation mode using warehousing robots, frees pickers from heavy work and greatly improves the efficiency of picking. With the development of intelligent warehouses, how to improve the operating efficiency of storage AGV has gradually become a major difficulty in the research and development of storage AGV.

On the one hand, the road network is the basis of AGV walking. Reasonable road network planning can effectively solve the problems of collision, deadlock, and low efficiency in the operation of AGV, thereby improving the operation efficiency of AGV and the efficiency of warehousing. The warehouse road network planning method is borrowed from the urban road network planning method, using diversion and collection as the basic planning concept, using one-way traffic organization to share the traffic pressure on the arterial road, and simplifying the intersection organization to improve the efficiency of AGV traffic. At present, the research on warehouse road network planning mostly adopts the method of

mathematical modeling. By constructing models and calculating the influencing factors and related indicators of the road network, the road network is planned and evaluated. For AGV road network planning, the solution algorithms mainly include ant colony algorithm [4], Dijkstra algorithm [4], A* algorithm [5] and genetic algorithm [6], but these algorithms can only solve the shortest path planning problem of a single AGV. When multiple AGVs are working at the same time, how to effectively avoid conflicts between AGVs is particularly important. For the multi-AGV scheduling problem, Chawlal et al. [7] compared five dynamic scheduling rules. Fanjupeyro et al. [8] proposed precise algorithm, because Multi-AGV scheduling problem is equivalent to Unrelated Parallel Machine Scheduling Problem (UPMSP). Zhang et al. [9] used the improved A* algorithm reservation table and dynamic weighted map to realize multi-AGV dynamic coordinated road network planning. Malopolski [10] proposed a method to prevent multiple AGVs conflicts and deadlocks based on the retention chain for a transport system with a square topology. However, there are few studies on road network planning using simulation methods. Aiming at the multi-AGV environment of the automated sorting intelligent warehouse for e-commerce, this paper focus on the comprehensive optimization of multiple AGVs scheduling and road network planning. Moreover, this paper uses simulation as the main method to study the road network planning of multiple AGVs operation.

On the other hand, whether in the traditional human-to-goods picking warehousing system or the latest goods-to-human intelligent warehousing system, there are not many types and quantities of items contained in a single order. In order to improve the efficiency of order picking, the distribution center usually combines several orders into a pick list for picking. Order batching is to divide the orders in the order pool into several batches according to certain rules. Each batch of orders constitutes a picking list and is picked in one operation. In an AGV-based intelligent warehouse, adopting a reasonable order batching strategy can greatly reduce the number of picking operations by staff and the number of times that warehouse robots move shelves, reduce picking costs and improve picking efficiency.

Based on the least time, Kang et al. [11] used a real-time task assignment method on the AGV transportation tasks with automatic assignment when the urgency of outbound goods is changing dynamically. Nicolas et al. [12] conducted experiments in an automated warehouse with multiple vertical lifting modules, established an optimization model with the shortest time to complete order picking, and designed a meta-heuristic method to solve the model. Xiang et al. [13] studied the storage allocation and order batching problems in the Kiva system, and minimized the number of shelf visits as the goal of order batching. The initial feasible solution is obtained by maximizing sequence association or minimizing sequence alienation, and the feasible solution is improved by the variable neighborhood search algorithm. Most of the research on the order batching problem is clustering based on the order similarity. In the traditional human-to-goods picking warehousing system, the number of pickers crossing aisles is the main factor affecting picking efficiency, so the order similarity is mainly defined according to whether the items in the order are located in the same aisle. In the AGV-based intelligent warehouse system, the

number of times the robot moves the shelf and the number of times the picker picks items from the shelf are the two main factors that affect the efficiency of order picking. Most of the existing literature defines order similarity based on one of the factors. Research result that considers both factors was not found.

This paper will also study the problem of order picking in the AGV-based intelligent warehouse system. Under the premise of a given warehouse layout and location allocation information, the product information contained in the order and the shelf information of the product in the order are also considered to construct a weighted order similarity. Moreover, the order batching strategy will be studied based on this indicator, through establishing a mathematical model of the order batching problem and designing an algorithm to solve the model. At the same time, it considers the cost of picking goods from the shelves and the cost of AGV handling shelves, so that the total cost of order picking in the intelligent warehouse system is minimized.

In Sect. 2, this paper will build an intelligent storage simulation model in Anylogic simulation software to count the number of collisions, path congestion, task completion time and other data in AGV operation under different road networks, and continuously optimize the road network by changing the warehouse layout, etc. In Sect. 3, in order to cope with the rapid increase in the number of orders, this paper builds an AGV order batching integer planning model. Section 4 is the conclusion, that is: through applying the optimized intelligent warehousing AGV road network, combined with the integer programming model and algorithm for order batching, e-commerce intelligent warehouse can effectively increase the efficiency of AGV order picking, reduce AGV operating costs, therefore improve e-commerce logistics service quality and customer satisfaction.

2 Optimization of Warehouse AGV Network Based on Simulation

2.1 Intelligent Warehousing AGV System

With the rapid development of informatization and intelligence, the involvement of intelligent warehousing AGV in logistics operations has gradually deepened, and are gradually replacing traditional manual operations and semi-mechanized logistics operations. Intelligent warehousing AGV can meet the current logistics needs with a large variety of products and large orders, and easily achieve the logistics operation goal of efficient, accurate, and timely. Therefore, various intelligent warehousing AGV systems have become a hot spot in the logistics industry. From the perspective of the functions implemented by the robots used in the logistics field, they mainly include handling, unstacking, palletizing, picking, sorting and other

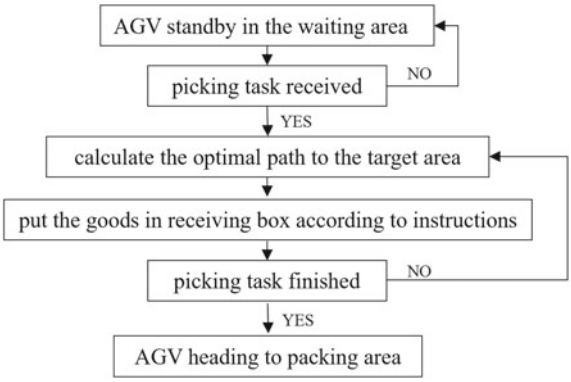


Fig. 1 Intelligent warehousing AGV operation flow chart

operations. Take the Feima robot in the Asia No. 1 smart warehouse of JD.com, China's second largest e-commerce platform, as an example, the operation flow is shown in Fig. 1. In this AGV system, the picker only needs to wait in his packing area without walking back and forth. This shortens the manual walking route, improves the efficiency of the warehouse, and reduces labor costs.

2.2 Intelligent Warehousing AGV Road Network Planning

Intelligent warehousing road network planning is to increase the traffic capacity in the designated area and reduce the probability or degree of congestion by dividing the road levels in the area, setting the traffic rules, and standardizing the walking of vehicles. It is effective mean to improve the operating efficiency of the intelligent warehouse. Whether the road network planning is reasonable or not is related to the traffic order and traffic operation efficiency in the intelligent warehouse. The road network of intelligent warehousing can impose corresponding constraints on the direction, speed, and driving road of the AGV, so that the operation of the intelligent warehousing becomes efficient and orderly. Through the planning of the intelligent storage road network, collisions, deadlocks and local or even large-area congestion problems can be avoided, and finally the stability, efficiency and reliability of AGV system in a complex and changeable environment can be realized. Road network planning is the basis for subsequent warehouse simulation work. From the perspective of logistics simulation, the ultimate goal of most simulation models is to find the best layout of smart warehouses and the best resource allocation schemes for efficient operation. The intelligent storage road network serves as the basis for the subsequent AGV walking, and its planning is reasonable or not has a direct impact on the subsequent simulation results.

2.3 AGV Road Network Simulation Model Based on Anylogic

Anylogic is a widely used tool software that can model and simulate discrete, system dynamics, and hybrid systems. Anylogic uses active objects to simulate different things in the real world. Active objects can use parameters and variables to represent the attributes of things, which are suitable for simulating fixed resources in the real world, such as shelves, AGVs, etc. At the same time, the simulation model built by Anylogic has the characteristics of visualization. Therefore, Anylogic can be used to simulate the entire process of the AGV from accepting the picking task to the end of all tasks, and extract information such as the number of AGV collisions, congested road sections, and the completion time of all picking tasks. According to the obvious operational problems in the AGV simulation process, combined with the data information extracted in the simulation to guide the optimization of the road network. The road network design in the simulation model includes: first, the statistical function of the number of collisions. AGV will collide during operation, and the number of collisions is one of the important indicators to measure the superiority of the road network. Using the same resources on different road networks to complete the same task, the number of final AGV collisions will also vary greatly. Reasonable road network planning can minimize the number of collisions of AGVs, thereby improving the operating efficiency of vehicles.

The calculation method of the number of collisions c of n AGVs in the simulation is as follows:

$$x_i = \sum_{j=1}^{n} y_{ij} \tag{1}$$

$$c = \sum_{i=1}^{n} \sum_{j=1}^{n} y_{ij} \tag{2}$$

$y_{ij} = 0$ (the distance between the j-th AGV and the i-th AGV is less than d),
$y_{ij} = 1$ (the distance between the j-th AGV and the i-th AGV is greater than d).

Among them, the d is affected by factors such as the size, shape, turning radius of the AGV, the AGV's own settings, and the actual operation of the intelligent warehouse.

Second, path popularity statistics. In the simulation model, the number of AGV passing through a specific path in a period of time is used as the path popularity of this section. Path popularity refers to the proportion of the number of times that the AGV passes through all the road segments in a specific road segment. Path popularity can reflect the degree of congestion of the path to a certain extent, and reasonable road network planning can allocate AGV traffic to the greatest extent to reduce congestion and improve AGV operation efficiency.

Third, the task completion time statistics. For a same task, the shorter the time it takes to complete, the higher the efficiency of the intelligent warehouse operation and the more reasonable the road network planning. The length of time to complete the simulation task is one of the main indicators to measure the superiority of the road network.

Table 1 is the basic parameters of the warehouse used in the road network optimization simulation experiment, which are typical storage warehouse parameters in the logistics field.

Figure 2 shows the warehouse simulation model and the visualization of the simulation results. The operation efficiency of AGV is restricted by many aspects. For example, the layout of the warehouse, the allocation of warehouse resources, and the work efficiency of personnel will all become bottlenecks that affect the efficiency of intelligent warehousing. Under a specific warehouse layout, the location and number of road network nodes and the setting of packing stations will affect the operating efficiency of intelligent warehouses. In order to reasonably plan the road network, so as to achieve better operating results, the overall road network can be adjusted according to the results of the simulation and the problems that appear in the simulation. By adjusting the road network under different schemes, several sets of experiments were carried out and the final simulation data were

Table 1 Parameters of the warehouse used in the road network optimization simulation experiment

	Warehouse related parameters	Specification
1	Size of warehouse	41 × 80 m
2	Path width	2 m
3	Size of shelf	10 × 0.5 × 2 m
4	Number of shelves	68
5	Time for picking goods per personnel	8.7 s
6	Time for personnel to load and unload baskets	10 s

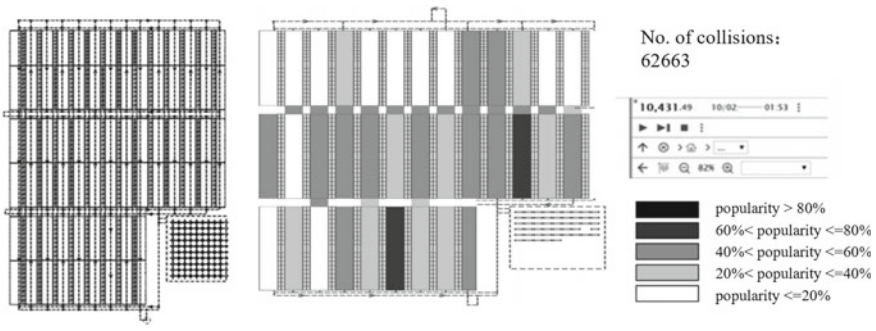


Fig. 2 Warehouse simulation model and visualization of simulation results

Table 2 Simulation data for different scenarios (part of data)

No. of AGV	No. of packing area	Experiment option	No. of collisions	completion time (s)	No. of area popularity >60%
50	1	1	65,656	10,431.49	5
	1	2	56,324	11,700.50	4
	2	3	70,433	12,455.00	2
	2	4	63,265	11,186.33	7
	2	5	62,774	11,998.40	3
	2	6	62,663	10,976.50	2
	3	7	78,980	9035.40	1
	3	8	64,877	10,924.60	1
	3	9	67,783	10,760.43	2
	3	10	58,966	12,519.09	3

recorded. The evaluation criteria of the road network are mainly measured by three indicators, the number of collisions, the completion time, and the area where the path popularity is greater than 60%. Some of the simulation data are shown in Table 2. By comparing the number of collisions, task completion time, and popularity, a more reasonable road network planner was finally determined as option 6, which layout is shown in the middle of Fig. 2.

3 Order Batching Model of Intelligent Warehouse AGV System

Based on the warehouse layout and location allocation information obtained from the aforementioned simulation experiments, the next step will focus on AGV order picking in the intelligent warehouse system. The order batching strategy will be explored based on the product information contained in the order and the shelf information of the product in the order, so that the efficiency of the intelligent warehouse system to complete order picking can be further improved, while the total cost can be minimized, and moreover, the customer service level and customer satisfaction can be improved.

3.1 AGV Order Batching Model Hypothesis Description and Analysis

If there are S shelves in the smart warehouse, each shelf has h storage spaces, and each storage space can store at most one kind of commodity. There is a total of

M kinds of commodities in the warehouse, and the storage space of each commodity is known. Suppose that there are N orders in the order pool at a certain moment in the warehouse that need to be picked. The product item information contained in each order, the cost of picking a product from the shelf, and the cost of AGV transportation are known. How to divide N orders into batches to minimize the total cost of order picking?

In order to simplify the problem, it is assumed that the goods in each order are not out of stock; the position of each product on the shelf is fixed, and the position of each shelf in the warehouse is fixed. AGV is required to select each batch of orders. The baskets of this batch of order goods are moved to the packing area, waiting for the staff to remove the goods and picked. When picking the next batch order, AGV needs move again, that is, the same kind of goods in different batch orders cannot be picked together.

Under the above assumptions, the main factors affecting the picking efficiency and picking cost of each batch of orders include two aspects. One is the number of times the staff picks the goods from the shelf. The other is the number of times the AGV moves the goods to the packing area and walking distance back and forth. Since the position of each shelf in the warehouse is determined, in order to simplify the problem, the transportation distance can be ignored when calculating the order picking cost, but only the times of transportations. Based on the above description, when the order is divided into batches, the product item information contained in the order and the shelf information of the products contained in the order should be considered comprehensively.

3.2 Integer Programming Model of AGV Order Batching

In order to establish the mathematical model of AGV order batching, the following symbols are defined (Table 3).

$$\min f = c_1 \sum_{k=1}^{K} \sum_{t=1}^{M} z_{kt} + c_2 \sum_{k=1}^{K} \sum_{s=1}^{S} y_{\mathrm{ks}} \tag{3}$$

$$\text{s.t.} \sum_{k=1}^{K} x_{\mathrm{ik}} = 1, \forall i \tag{4}$$

$$\sum_{i=1}^{N} x_{\mathrm{ik}} \leq \mathrm{q}, \forall k \tag{5}$$

Table 3 Symbols in mathematical model of AGV order batching

Index	Parameters	Decision variable
i, j: order index, $i, j = 1, 2, \ldots, N$	q: the maximum number of orders allowed for each batch	$x_{ik} = \begin{cases} 1, & \text{order } i \text{ is allocated to batch } k \\ 0, & \text{otherwise} \end{cases}$
k: batch index, $k = 1, 2, \ldots, K$	$a_{it} = \begin{cases} 1, & \text{the order } i \text{ contains goods } t \\ 0, & \text{otherwise} \end{cases}$	$y_{ks} = \begin{cases} 1, & \text{picking batch } k \text{ needs to carry rack } s \\ 0, & \text{otherwise} \end{cases}$
s: shelf index, $s = 1, 2, \ldots, S$	$b_{ts} = \begin{cases} 1, & \text{commodity } t \text{ is stored on shelf } s \\ 0, & \text{otherwise} \end{cases}$	$z_{kt} = \begin{cases} 1, & \text{batch } k \text{ contains commodity } t \\ 0, & \text{otherwise} \end{cases}$
t: commodity index, $t = 1, 2, \ldots, M$	c_1: the cost of picking a commodity from the shelf	
	c_2: the cost of per AGV moving	

$$\sum_{i=1}^{N} a_{\mathrm{it}} x_{\mathrm{ik}} \leq \mathrm{N} \cdot z_{kt}, \forall k, t \tag{6}$$

$$z_{kt} \leq \sum_{s=1}^{S} b_{\mathrm{ts}} y_{\mathrm{ks}} \forall k, t \tag{7}$$

$$x_{\mathrm{ik}} \in \{0, 1\}, \forall i, k \tag{8}$$

$$y_{\mathrm{ks}} \in \{0, 1\}, \forall k, s \tag{9}$$

$$z_{kt} \in \{0, 1\}, \forall k, t \tag{10}$$

The objective function (3) represents the total cost of minimizing order picking. The first term represents the cost of picking goods from the shelf, and the second term represents the cost of handling. The constraint (4) means that each order is allocated to a batch. The constraint condition (5) means that the number of orders allocated to each batch does not exceed the specified maximum quantity. The constraint condition (6) means that if any order in batch k contains product t. The batch constraint condition (7) means that if batch k contains commodity t, the shelf containing commodity t needs to be transported when picking this batch. Constraint conditions (8–10) represent the value constraints of decision variables.

The theorem: AGV order batching of intelligent warehouse system is an NP-hard problem. Since the order batching problem is an NP-hard problem, small-scale problems can directly use commercial solvers such as Lingo, Cplex, etc. to solve the integer programming model and obtain the optimal solution. Large-scale problems need to design fast and effective approximate solving algorithms.

According to the objective function and constraint conditions of the order batch mathematical model, the main factors affecting the order picking cost are the number of items contained in each batch and the number of shelves that meet the picking requirements of each batch. Therefore, merging orders that contain the same product items or that require the same shelf can effectively reduce the total cost of order picking. Therefore, the indicators describing the degree of order similarity are whether the items included in the order are the same and whether the shelves of the items included in the order are the same. Based on the order similarity of items and shelf-based order similarity, the weighted similarity can be further defined as the basis for order batching.

Since e-commerce intelligent warehouse involves a huge number of orders, items and shelves, the corresponding order batching problem is complicated, and it is impossible to obtain an accurate solution by directly solving the integer programming model in a short time. Hence it needs to design a fast and effective approximate solution algorithm to obtain an approximate optimal solution, and can use a greedy algorithm based on the weighted similarity between orders to solve the order batching. In detail, first, calculate the weighted similarity between the orders

according to the items contained in the order to be picked and the shelf information of the various products and treat each order as a single batch. Then calculate the average of any two batches after merging the amount of weighted similarity increases, and the batches that meet the total number of orders in the batch after the merger. The maximum increase in the average weighted similarity increase are selected for merging, until the average weighted similarity no longer increases or there is no longer a merge to meet the capacity constraints order batches.

4 Conclusion

This paper takes e-commerce intelligent warehouse AGV as the research object. Through the optimization of intelligent warehouse AGV road network and order batching, the research goal is to achieve the highest operating efficiency and the lowest warehousing cost, so as to increase e-commerce customer service level, and improve e-commerce customer satisfaction. Firstly, the use of reasonable planning of warehouse road network can effectively solve the problems of collision, deadlock, and low operating efficiency that occur during the operation of AGV. This paper uses Anylogic simulation software to build an intelligent storage simulation model to simulate the AGV operating, path popularity, number of collisions, and task completion time under different conditions. Through the simulation process, the warehouse road network is continuously optimized, and the warehouse road network planning scheme that makes the AGV more efficient is determined. Secondly, due to the continuous increase in the number of e-commerce orders, based on the optimization of the AGV operation road network, this paper also studies the batch picking, through analyzing the two main factors affecting order picking efficiency and picking costs in intelligent warehouses. The factors are the number of times the staff picks goods from the shelves and the number of times the AGV moves the shelves. With the goal of minimizing the total cost of order picking, an integer planning model of order batching is established. According to the product information contained in the order and the product correspondence, order similarity index is constructed based on the shelf information, and a greedy algorithm based on the weighted similarity between orders is designed. In the follow-up research, specific calculation examples will be used to perform simulation calculations, so that the research method in this paper can not only suitable for solving AGV order batching in intelligent warehouses, but also for solving other types of unmanned warehouse systems in e-commerce, such as unmanned warehouse based on rotating shelves and conveyor belts, etc.

References

1. Ministry of Communications and Multimedia (KKMM), Malaysia (2020) (Online). Available from: https://www.kkmm.gov.my/. Accessed 20 Nov 2020
2. Logistics order statistics for double eleven (11.11) from 2009 to 2020 (2020) (Online). Available from: https://weibo.com/tv/show/1034:4570441583624244?from=old_pc_videoshow. Accessed 13 Nov 2020
3. Bogue R (2016) Growth in e-commerce boosts innovation in the warehouse robot market. Ind Robot Int J 43(6):583–587
4. Baeza D, Ihle CF, Ortiz JM (2017) A comparison between ACO and Dijkstra algorithms for optimal ore concentrate pipeline routing. J Clean Prod 144:149–160
5. Duchon F, Babinec A, Kajan M (2014) Path planning with modified a star algorithm for a mobile robot. Proc Eng 96:59–69
6. Liu E, Yao X, Lan H (2018) AGV dynamic path planning based on improved genetic algorithm and its implementation. Comput Integr Manuf Syst 24(6):1455–1467
7. Chawlal V, Chanda A, Angra S (2018) Simultaneous dispatching and scheduling of multi-load AGVs in FMS-A simulation study. Mater Today: Proc 5(11):25358–25367
8. Fanjupeyro L, Ruiz R, Perea F (2019) Reformulations and an exact algorithm for unrelated parallel machine scheduling problem with setup times. Comput Oper Res 101:173–182
9. Zhang D, Sun X, Fu S (2018) Cooperative path planning in multi-robots intelligent warehouse. Comput Integr Manuf Syst 24(2):410–418
10. Malopolski W (2018) A sustainable and conflict-free operation of AGVs in a square topology. Comput Ind Eng 126:472–481
11. Kang Y, Feng G, Wang Z (2020) Real-time task assignment method of two-load AGV under dynamic change of goods urgency in logistics warehouse. J Phys: Conf Ser 1576(1):12–55
12. Nicolas L, Yannick F, Ramzi H (2018) Order batching in an automated warehouse with several vertical lift modules: optimization and experiments with real data. Eur J Oper Res 267 (3):958–976
13. Xiang X, Liu C, Miao L (2018) Storage assignment and order batching problem in Kiva mobile fulfilment system. Eng Optim 4:1–22

Analyzing Tweet Data to Identify the Impact of Demonetarization

Kannan Suseelan Unnithan, Said A. Salloum, and Khaled Shaalan

Abstract In this century, we have witnessed tremendous growth in the application of technology in our day-to-day life. One such life influencing factor is social media. In a manner or other, each individual's life is influenced by social media. Individuals and organizations rely more on social media to interact with people, acquire knowledge, promote business, gather information, news telecast, and for many other professional and personal purposes. Researchers use social media as one of the primary data sources for studies related to people's behavior and for various prediction purposes. This empirical study identifies the importance of social media data in recognizing any event's influence on the public. In October 2016 Indian government had demonetarized some active currency, and it created mixed reactions in public. For this research, tweet data was used to determine the impact of demonetarization on India's people. The study results demonstrate that tweet data can be relied on identifying or extracting general public opinions on any given topic or event.

Keywords NLTK · Tweet extraction · Demonetarization

1 Introduction

An online platform where people form networks with other individuals or groups of people with similar professional or personal interests, background, activities, or real-life connection is termed as Social Media [1–5]. Through social media, people

K. S. Unnithan · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_7

share their ideas, videos, and photos and discuss any subject of interest with their connections publicly or privately [6–10]. In this era of technology, people are heavily dependent on the internet to interact with others or keep up-to-date with the happenings around them [11]. Individuals express their opinion or feedback about any business entity or social engagements through the internet. Investors rely deeply on the internet to get the opinion from others [12].

Sharing information is the main objective of any social network, and this characteristic made them popular [13]. The popularity of social media made them one of the largest databases of human information. The information in social media includes the experience of individuals, people relations, feelings, national and international news, etc. This information can be used to analyze human personality, trend predicting for present and future events, mining user's personal states, and emotion analysis [14]. As the web and social media information grow, the main concern is about extracting beneficial data or generating useful reports from those.

As the use of social media increases, the internet becomes a massive and reliable source of information for businesses and individuals. Social networks like Facebook, Twitter, LinkedIn; media sharing networks like Instagram, Snapchat, YouTube; Discussion forums like Reddit, Quora, Digg; and Blogging & publishing networks like WordPress, Tumblr, Medium are the main popular social media. Effective taping of social media data will help in areas like campaign planning, market demand study, social selling, trending topic identification, customer service, recruitment… etc.

For researchers, social media is a gold mine of data. Being one of the top social media networks, Twitter is considered as a major data source for researchers. Data on Twitter are very specific, and most of the users' tweets are available for the public. Twitter provides APIs which allow pulling data from Twitter based on any particular topic [15]. The objective of this research is to study the role of social media data in spotting public opinion and to identify the Indian citizens' approach towards the demonetarization implemented by the Indian government in October 2016.

2 Literature Review

In recent years many researchers have used social media data to conduct studies on various topics like market analysis, election result prediction, stock market prediction… etc. [16–18]. In this research, a detailed literature review was conducted to identify the application of social media data for professional and personal purposes.

Zhang et al. [19] argue that stock market prediction can be improved by extracting and analyzing quantitative features, sentiments, and events from social media and the web. To prove this, the authors collected data from China and Hong Kong share markets, along with the matching event data and sentiment division from social media. The proposed CMT model performed better than the comparing

model SVM. They conclude that the stock market movement can be predicted effectively using social media data with factorized low-rank matrices and completing the missing values in the sparse tensor.

An automatic reporting framework has been proposed by Zhang et al. [20] to effectively identify popular actions from related websites and then improve them using data extracted from a multi-model social network to produce concluding reports for the stakeholders. They implemented a web crawling tool to observe six news websites and used Twitter streaming API to crawl geotagged tweets. A total of 7700 news records and 1,396,745 tweets were collected for the study. This study proved that the data extracted from both the websites and Twitter could be utilized to generated meaningful reports. Also, it showed that the Twitter data was complementing the finding from the website crawling and were very useful. This result was evaluated and confirmed using a user study of a group of 20 subject matter experts.

Another study shows that Twitter data can be utilized to extract valuable insights into the multiple facets of any disease outbreak [21]. In this study, they proposed a classification approach to effectively classify the tweets using low-level lexical class-specific features. Tweets related to two epidemics, Ebola in 2014–2015 and MERS in 2014, were used in this study. After cleaning and preprocessing, a total of 400,000 messages were included in the dataset. The study extracted the information from tweets and classified the same into five categories, Symptom; Prevention; Disease transmission; treatment; and Dearth report. The system output was verified by human experts, and an agreement of 87% was received. This study clearly showed that the Twitter data could be extracted and processed to provide useful insights for situations like a disease outbreak.

Community identification is the stepping stone to understand the fundamental structural and semantic information in any network. Interconnected groups or communities can be identified based on the edge weight related to the number of times a tweet republished by Twitter users [22]. Abdelsadek et al. [22] introduced a new algorithm called Tribase to analyze and identify communities from Twitter data. By comparing the algorithm with the similar algorithms available on the LFR benchmark showed the proposed Tribase algorithm works well in identifying communities from the Twitter data.

To identify Twitter user's nationality, Bayat and Ucan [14] used KNIME, an open-source data analytics. Using KNIME, over 300,000 tweets of 757 Twitter users were downloaded. The implemented algorithm identified the user nationality using the nationality of the users' network. The overall accuracy of the proposed algorithm was 90%, and it proved that using Twitter data, one could identify the nationality of the users even if they have not explicitly specified their country. Hoang and Mothe [23] also conducted a study in identifying location from the Twitter data. They identified that the Ritter tool identified locations combined with the locations identified with DBpedia's Stanford Filter increases the F-measure of the location extraction.

Social media data are being used in investigating human activity patterns in the urban area. Twitter data collected from the city of Los Angels during the time

period of Feb 2010 to Jan 2011 was used to extract human activity information to conduct the spatial analysis of crashes in the urban area [24]. A web crawler application developed in Python was used to collect the location type from the check-in data of Twitter. This study showed that the human activity variables in the Twitter data do not have any biased estimates for the traditional traffic exposure variables' influence. Twitter data can also be extracted to identify and present future events [25]. The event prediction system used clustering techniques based on the users' interest and event start and end time. The evaluation of the algorithm gave 84% of accuracy.

All the previous studies show that social media data, especially Twitter data, can be utilized to study or predict the characteristics of any type of business, and can be used for sentiment analysis, market trend analysis, stock market prediction, human activity analysis, event prediction, community identification, etc. In most of the reviewed literature, Python, NTLK was used for downloading Twitter data and processing the tweets.

3 Research Methodology

This study used the tweet data from Twitter to identify the public reaction to the demonetarization implemented in India. For this purpose, a data set of 3 months tweet data, with the hashtag #demonetisation, from 8 Nov 2016 was used. Twitter developer account was used to extract the data set from Twitter using the Python/ NTLK.

The data extraction process had two parts:

1. Access and download Twitter Data
2. Information extraction from the Twitter data.

Unlike any other social media network, most of the Twitter data are public and can be accessed by anyone. However, to access Twitter data, one has to register a developer account with Twitter to get access keys. Twitter developer application to generate authentication keys can be submitted through the URL https://apps.twitter.com.

After the successful registration, Twitter will provide the following access keys to use the Twitter APIs.

- CONSUMER_KEY
- CONSUMER_SECRET
- ACCESS_TOKEN
- ACCESS_TOKEN_SECRET.

OAuthHandler function available in the Tweepy package was used to access Twitter. Using the Cursor function of the Tweepy, Twitter data related to the demonetization is downloaded to a CSV file. This data set included 14,940 tweet data related to demonetization.

The downloaded dataset had many fields like 'text' which contains the tweet text; favorited, which specifies favorite status; favoriteCount, to show the count of favorite; replyToSN, who replied; created date, … etc. However, for this study's purpose, only field 'text', which contained all tweets' text, was considered.

After collecting the research data, the first step in information mining from unstructured text data is to clean the text or perform preprocessing. For this study, Python NTLK's 'stopwords' were used to preprocess the data. After the preprocessing, a new field named 'cleaned' was created. After getting the cleaned data, the next step is to analyze the same. The main types of information that can be extracted from tweet data are Keywords, Phrases, People Mentioned, Hashtags, URLs/Emails, Person Names, Company Names, and Sentiments. These are very high-level analysis, which was utilized for this research.

Now the Keywords have to be extracted from the newly created field 'cleaned', and then use the tokenizing technique to get individual words. The split function was used to separate the words in the joined text and used the counter function to get each word count and list the top 100 words. Thus, a list of the most common words used in the tweets related to demonetarization was generated. From this list, the most frequent words used by the users were identified. Further analysis can be done by selecting the tweets with these keywords.

In addition to the keywords, 'mentioned persons', hashtags, and links were identified for this study. The mentioned persons like celebrities or politicians were captured using the identification of the special character @. Similarly, the hashtags and links were extracted using identifying the words starting with '#' or 'http//', respectively.

Also, NLTK's ngrams package and Named Entity Recognition were used to analyze bigram, trigram, or ngram words and to extract the person or company name from the tweet dataset. This helped to extract and identify words/name like "Narendra Modi", "PM Narendra Modi", "Happy with demonetization", "foolish decision", "CorruptionFreeIndia" … etc.

4 Result and Conclusion

Social media is a huge source of data where much useful information is hidden. By managing and analyzing properly, many unforeseen information can be tapped from social media data, and Twitter is known as one of the ten most famous social media [14]. This research accessed data from one of the main social media Twitter and conducted the analysis of tweet data related to demonetarization executed in India during October 2016. The main analysis done through this study is based on keywords, mentioned persons, hashtags, links, ngram words, and Named Entity Recognition. By extracting and analyzing the tweet data using the Keywords, HashTags, and Mentioned Persons, it was observed that the public had a mixed reaction towards the Indian governments' demonetarization. The major drawback of this research is that the analysis was of a very high level. A detailed and specific

characteristic of the data was not identified. Also, a suitable validation mechanism should be applied. In the future, this research can be extended to: Unwrap the extracted shorten web links using some available Python packages, implement a detailed sentiment analysis by implementing sentiment classifiers using a machine learning classification model or use the 'textblob' library to implement basic sentiment analysis, conduct descriptive statistics to identify which group was supporting or opposing the demonetarization or identify the attitude of the organization towards the demonetarization and conduct time series analysis to identify how people opinion on demonetarization change as time changed.

References

1. Salloum SA, Al-Emran M, Monem AA, Shaalan K (2017) A survey of text mining in social media: Facebook and Twitter perspectives. Adv Sci Technol Eng Syst J 2(1):127–133
2. Mhamdi C, Al-Emran M, Salloum SA (2018) Text mining and analytics: a case study from news channels posts on Facebook. https://doi.org/10.1007/978-3-319-67056-0_19
3. Salloum SA, Mhamdi C, Al-Emran M, Shaalan K (2017) Analysis and classification of Arabic newspapers' Facebook pages using text mining techniques. Int J Inf Technol Lang Stud 1(2):8–17
4. Salloum SA, Al-Emran M, Shaalan K (2017) Mining text in news channels: a case study from Facebook. Int J Inf Technol Lang Stud 1(1):1–9
5. Salloum SA, Al-Emran M, Abdallah S, Shaalan K (2017) Analyzing the Arab Gulf newspapers using text mining techniques. In: International conference on advanced intelligent systems and informatics, pp 396–405. Springer. https://doi.org/10.1007/978-3-319-64861-3_37
6. Wikipedia. Social networking service, Wikipedia
7. Salloum SA, Maqableh W, Mhamdi C, Al Kurdi B, Shaalan K (2018) Studying the social media adoption by university students in the United Arab Emirates. Int J Inf Technol Lang Stud 2(3):83–95
8. Habes M, Salloum SA, Alghizzawi M, Mhamdi C (2019) The relation between social media and students' academic performance in Jordan: YouTube perspective. In: International conference on advanced intelligent systems and informatics, pp 382–392. Springer
9. Alghizzawi M, Habes M, Salloum SA (2019) The relationship between digital media and marketing medical tourism destinations in Jordan: Facebook perspective. In: International conference on advanced intelligent systems and informatics, pp 438–448. Springer
10. Al-Maroof RS, Salloum SA, AlHamadand AQM, Shaalan K (2019) A unified model for the use and acceptance of stickers in social media messaging. In: International conference on advanced intelligent systems and informatics, pp 370–381. Springer
11. Almazrouei FA, Alshurideh M, Al Kurdi B, Salloum SA (2020) Social media impact on business: a systematic review. In: International conference on advanced intelligent systems and informatics, pp 697–707. Springer
12. Drake MS, Thornock JR, Twedt BJ (2017) The internet as an information intermediary. Rev Account Stud 22(2):543–576
13. Foroozani A, Ebrahimi M (2019) Anomalous information diffusion in social networks: Twitter and Digg. Expert Syst Appl 134:249–266
14. Bayat O, Ucan ON (2018) Estimation of Twitter user's nationality based on friends and followers information. Comput Electr Eng 66:517–530
15. Young M (2017) Twitter data mining: a guide to big data analytics using Python
16. Al Mansoori S, Almansoori A, Alshamsi M, Salloum SA, Shaalan K. Suspicious activity detection of Twitter and Facebook using sentimental analysis

17. S.A. Salloum, "Predicting the Intention to Use Social Media Sites: A Hybrid SEM-Machine Learning Approach," Advanced Machine Learning Technologies and Applications: Proceedings of AMLTA 2021, 324.
18. Almansoori A, AlShamsi M, Salloum SA, Shaalan K (2021) Critical review of knowledge management in healthcare. 99–119. https://doi.org/10.1007/978-3-030-47411-9_6
19. Zhang X, Zhang Y, Wang S, Yao Y, Fang B, Philip SY (2018) Improving stock market prediction via heterogeneous information fusion. Knowl-Based Syst 143:236–247
20. Zhang J, Li X, Nie W, Su Y (2017) Automatic report generation based on multi-modal information. Multimedia Tools Appl 76(9):12005–12015
21. Rudra K, Sharma A, Ganguly N, Imran M (2018) Classifying and summarizing information from microblogs during epidemics. Inf Syst Front 20(5):933–948
22. Abdelsadek Y, Chelghoum K, Herrmann F, Kacem I, Otjacques B (2018) Community extraction and visualization in social networks applied to Twitter. Inf Sci 424:204–223
23. Hoang TBN, Mothe J (2018) Location extraction from tweets. Inf Process Manage 54(2): 129–144
24. Bao J, Liu P, Yu H, Xu C (2017) Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. Accid Anal Prev 106:358–369
25. Kunneman F, Van Den Bosch A (2016) Open-domain extraction of future events from Twitter. Nat Lang Eng 22(5):655–686

Qualitative Study in Natural Language Processing: Text Classification

Ahlam Wahdan, Said A. Salloum, and Khaled Shaalan

Abstract Recently Natural Language Processing (NLP) has excessive attention due to increased data available online. Although there is huge development in Arabic NLP, but still it is behind English NLP. The aim of this study is to examine the most useful qualitative research design that can be used in the field of Arabic Natural Language Processing in general and in-text classification in specific. Two research designs have been examined in detail; survey and experimental research. Both benefits and drawbacks illustrated for each research design with examples for each method. Thus to have better understanding for some of the research methods. Then deciding which one is best fit to design a qualitative study in this field. Furthermore, this paper suggested framework can be used while researching in the field of NLP. The framework suggested to start with survey or Literature review then follow it with experiment.

Keywords Arabic · Natural Language Processing (NLP) · Text classification

1 Introduction

In the current era, the rapid increase in the volume of the data published over the networks increased the need for new techniques to extract the knowledge, analyze, and summarize it. So artificial intelligence comes to resolve these issues by providing many domains that facilitate extraction summarization and analyzing for data [1–7]. One of these domains is Natural language processing, which specialized

A. Wahdan · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_8

in analyzing natural human language [8–11]. Text summarization, clustering, machine translation, sentiment analysis, and named entity recognition are some applications of applying natural language processing. Arabic natural language processing is very challenging because of complex structure and morphology; thus, the number of researches in Arabic NLP is not much as English NLP [12–15]. Many techniques have been used to improve the application of Natural Language Processing, such as word embedding, deep learning, statistical methods, and graph representation [16].

Text classification is the labeling or categorization process that aims to tag the text document and discriminate its type or category. Text can be categorized into a single label and can also be categorized into a more challenge type, which is a multi-label [17]. Many Major challenges are facing Arabic text classification that makes the number of researches still few regarding the researches published for English languages. Arabic language is complex, delict morphology; also, it is a rich language. A huge number of features representing the data is another major difficulty in Arabic text classification. Where having more features will increase the runtime. So reducing the number of features is a solution to overcome this issue and increase the performance. Thus feature selection methods can be applied in order to choose the features that are more representative of the data [18].

This paper is very important because it looked at a number of peer-reviewed articles in the NLP with the main objective of showing the most efficient text classification methods. Then we have suggested a new framework to be followed in qualitative studies for text classifications. This framework could be a guideline for the practitioners in the field of NLP.

The structure of this paper as follows: Sect. 2 summarizes the qualitative methods have been used in Arabic text classification. Section 3 illustrates pros and cons for each qualitative methods. Section 4 explain the suggested approach will be used in our study of Arabic text classification. Section 5 is the conclusion.

2 Qualitative Methodologies

This part summarizes the methodologies used in researches in the field of Arabic natural processing. Two research design methods have been declared; Survey and Experimental research with examples from published research papers on each method.

2.1 Survey

The survey in the NLP field is not typical; it is more likely to systematic and peers review for previous researches. The researcher will determine his survey questions

and start looking for their answers in the chosen paper. In the end, it is the researcher's judgment or opinion.

2.1.1 Survey 1

Title: "Graph-based Arabic NLP Techniques: A Survey"
Graph representation of Arabic Natural processing was the base of this study. Graph method is a type of categorization where it classifies the data into different categories using the nodes and edges of the graph itself. There are different data types that graphs can represent, such as networks, text components, and webpages. Also, there are different types of categorized graphs, such as directed/undirected graph, weighted/unweighted graphs, and cyclic/acyclic graphs. Mainly this study investigates ANLP major techniques in several researches related to graph theories. Also, it aims to discuss the problems facing natural language processing and the role of graph-based techniques in solving these problems. The methodology of this research was essentially conducting a survey by reviewing published researches to build a clear idea about the graph-based method as a representation for the state-of-the-art in this field. This survey targeted the researches that discussed the applications used in ANLP and the problem faced when graph method had been used. The survey conducted a review by formulating six main questions that serve the aim of this study. Besides, it covered 11 researches. The survey revealed a good comparison between the researches in the aspect of the types of graph used and the core technique in each study. Also, it explored the best approaches that lead to better results while using graph methods. Moreover, this survey concluded that ANLP still has significant challenges, especially in the resources; hence the researches still in its early stage, and there is a lack of the resources available. The survey clears up the gaps to be a base for future work, which is improving the performance of the current application in the field of ANLP by building a hybrid model that combines graph method with another technique such as neural networks and statistical method [16].

2.1.2 Survey 2

Title: "A Survey of Arabic Text Mining"
The aim of this study is to inspect the techniques used in Arabic text mining, especially non-structured Arabic text. Also, it aims to investigate the implementation of text mining in three different categories; the Holy Quran, sentiment analysis, and web documents. Text mining is extracting and retrieving knowledge from unstructured text. It can also be called Knowledge-Discovery in Text (KDT) or Intelligent Text Analysis. Text mining is very important; hence 80% of the data saved in text format, so potentially it has great value. The methodology of this research was essentially conducting a survey by reviewing published studies to build a clear idea about text mining and focusing on the applications and techniques

used. This survey targeted the researches that discussed the forms used in Arabic text mining, and the techniques had used with detailed explanation for the methodology of each. The survey conducted a review by formulating three main questions to focus on the problem, methodology, and the result. As mentioned previously, this survey divided text mining into three categories. The first one Text mining—Holy Quran, where only three researches covered. The second category; Sentiment Analysis, which includes eight researches. In the last category, Text Mining—Web documents, which includes 14 researches. The survey clears up the gaps that are highly required to be a base for future work, which is focusing more on analyzing online unstructured Arabic text such as social web data; hence it needs a lot of improvements. Also, this survey stated that sentiment analysis also needs more researches and development [19].

2.2 Experimental Research

2.2.1 Experiment 1

Title: "Automatic Arabic text categorization: A comprehensive comparative study" The aim of this study is to concentrate on text categorization or classification, which means classify the documents into different categories regarding the content of the text documents. This process is very crucial because of the rapid increase in the online texts published daily on the internet, where manual classification becomes impossible. This paper compared text classification algorithms. Also, it discussed how the efficiency of classification techniques would be affected by changing Arabic stemmers in both types; light and root based stemmer.

The methodology of this paper is experimental, where the author experimented with two data mining software, which is Weka and Rapid minor. Initially, the author defined the corpus he used. It is open-source corpus "Arabic articles," which contains 2700 documents and includes nine categories. The experiment has been done in 3 main steps started with preprocessing, processing, test, and evaluation. The preprocessing step includes removing stop world, tokenization, weighting. The preprocessing step has been done in both Rapidminer and Weka. In the processing step, five classifiers have been applied; Naïve Bayes, SVM, KNN, Decision tree, and Decision table. The corpus had been divided into two parts training and testing set—the testing set used for testing and evaluation to test the models and evaluate their efficiency. Cross-validation with 5-folds had been used, and the efficiency measures where accuracy, precision, recall, and F1 score. The author compared Weka and rapid miner results using the same dataset and same algorithm each time. The results regarding the best classifier were Support vector machine. Regarding the preprocessing stage, the stemmer that outperforms well was light ten stemmers. Finally, the author recommends using Rapidminer; hence its scalability and efficiency exceed Weka [20].

2.2.2 Experiment 2

Title: "Arabic text classification using deep learning models"
This study designed to focus on using deep learning technique for Arabic Text Classification. So several models have been built and compared. Nevertheless, all of the models where types of deep learning as mentioned before. In this experiment, the text had been categorized and labeled into two types, single and multi-label. This study differentiated itself from other researches by three remarkable things. Firstly, the researcher has built two huge corpora and make it available online for any future research; SANAD and NADIA. The first one has been used for single labeling and the second used for multi-labeling classification. Secondly, it doesn't include normalization. Finally, this study focused on improving classification performance by using the word2vector model. Also, the researchers benchmark their own dataset with another dataset. The methodology that has been used is experimental. The researchers have introduced two corpora; SANAD and NADIA. Many models have been built, but all of them were based on convolutional and recurrent neural networks CNN and RNN. In the methodology, three main steps applied. The first step is the preprocessing step, which includes cleaning the data by removing spaces, punctuation marks, and removing non-Arabic characters. However, this study doesn't include normalization as the adopted practices. Because the researcher claimed that this would change the contextual meaning for some words, benchmarking with others in a single label and multi-label. Hence many sets of experiments have been done, but the first set of experiments involved comparing the results of this research corpora with the benchmark dataset. In the Implementation step, the data divided into three folds; 80% for training the models and 20% divided equally between validation and testing. Two frameworks have been chosen for the implementation, which is Keras and Tensorflow.

2.2.3 Experiment 3

Title: "An improved Chi-Square feature selection for Arabic text classification using decision tree"
The number of attributes used in text classification is huge, and there are many methods to reduce this number and choose the optimal number of attributes to be included in the model. So the aim of this study is to select the optimal number of features to reduce the runtime and to increase efficiency. The methodology used in this study is a comparative experiment, where it compares two methods of feature selection—Chi-square and ImpCHI. This study used a corpus with 250 documents and five categories. The evaluation for efficiency is a recall. The methodology of this study divided into three steps starting with a Literature review, then preprocessing step. They were followed by feature selection. Lastly, the analysis and result part where The Data has been divided into training with 150 document and testing with 100 documents. The preprocessing is the cleansing process, which includes removing unwanted characters such as digits, non-Arabic characters, and

stop words—followed by a normalization step. The second step is Stemming; were the words brought back to its root by removing prefixes, suffixes, and infixes. Third step; feature selection where two methods compared, Chi-square and improved chi-square (IMPCHI). Evaluating CHI square and improved CHI square has been done using light stemming and hard stemming. Moreover, the evaluation criteria have done using Recall measure. The result shows that IMPCHI outperforms Chi-square in classifying Arabic text, and it revealed the highest recall measure [18].

2.2.4 Experiment 4

Title: "Using Natural Language Processing and Qualitative Analysis to Intervene in Gang Violence: A Collaboration between Social Work Researchers and Data Scientists"
This study aims to reduce and prevent social media violence by developing a system that decodes social media posts from Twitter using natural language processing and qualitative analysis. The methodology of this research was essentially to use in-depth qualitative methods accompanied by coding for each post in social media. For the purpose of understanding the communication trends in the social media violence gang where the gang youth involved. Regarding the massive amount of data that will be collected, efficient and automatic coding will be needed. Moreover, these posts will include emoji's, images, and linguistic style. Thus natural language processing will be used. The study targeted Twitter communications of a female gang member from Chicago using Radian6 as a social media tracking system. Training data set have been created in a place where a high rate of the violent atmosphere [21].

3 Pros and Cons

In this section, we will make a comparison between the qualitative methods used in the research design of the previous studies. The qualitative methodology used in the previous studies is; surveys and experiments. Each researcher of these studies decided his own research methodology in order to resolve research questions and to derive the research objectives. This section will focus on its advantages and disadvantages.

3.1 Survey

Surveys are one of the qualitative methods used in research design. There are many types of surveys; it can be explanatory, exploratory, and also can be descriptive.

There are many ways to conduct the surveys; we have online surveys, phone-based surveys, interviews (face to face), and mail surveys. Each of them has its own pros and cons as well [22]. Generally, using surveys in research methodology have many advantages and disadvantages. Surveys have the advantage that it can collect data from a large sample of the population by taking care of choosing the population. More data give more insight into the trends after analyzing it. In general, the cost of implementing surveys is less than other qualitative methods. Also, the ease of implementing it is another advantage. Moreover, surveys can have a wider range of information about a specific topic; hence it has the ability to ask various questions. Nevertheless, each type of survey has its own advantage s and disadvantages [22].

Using mail to develop a survey has many advantages, such as the ease of implementing it and low cost. Also, people would be more open to sharing their information. On the other hand, mail survey has disadvantages; such as the response rate could be low; hence a lot of people will not respond, and also some people who are from low literacy cannot respond. Also, it is difficult to go back to the respondents and probe them. Using the phone as another type to do the survey where it could cover large-scale of population and could reach different countries. Also, the quality could be more controlled and could give more flexibility for the survey. But also there are some cons such as limitations in time and in open-ended questions. Also, the visual material cannot be valid; hence it is on the phone. Regarding the online survey, it has its own pros and cons—some of the pros similar to the mail type regarding the cost and to sharing information willingness. And moreover, and it is automated and needs less time. In regard to the cons, some are like a survey in the difficulties to go back to the respondents and probe them, and the response rate could be below. Face to face surveys can give a clearer attitude. Hence it can be observed, and the response time will be high, but the drawback of using this type is in the cost and time. Hence it is more costly than others and needs more time. Also, the result at the end may not represent the population [22].

3.2 Experimental Research

Some researches need to be tested so it can be generalized, especially if the variables can be controlled over the set of variables. There are many advantages of using experimental research methods in the research, having control over the variables just as mentioned before. This can be considered from the advantages because the researcher can isolate and determine specific factors. Nevertheless, it's basic; it is an efficient research method. Because it can be used in different scenarios. Moreover, in the analysis section, where the researcher can control the variable, the revealed outcome is more relevant than any other research method. The flexibility of the experimental research gives wide variation. Hence, the researcher can experiment with any unique situation and check out the outcome. These are the advantages of using an experimental research method in the research. But also there are drawbacks. Any type of error happens during the experiment

could completely destroy the validity of the experiment. Errors have many types; systematic error- which can occur during the experiment, or random error, which is uncontrolled or even human error. Another drawback is artificial solutions can be made. This is because of the over-controlled variable, which will cause skewness or corrupting for the data. Also, experimental research needs to test each variable in the experiment, and some experiment needs to test all the probabilities of the combined variables, which is time-consuming, resource and finance consuming as well [23].

4 Preferred Method

In general, the methods used to build a qualitative search are: Simulation, Experiment, Survey or Literature review, as clear in Fig. 1.

In-text classification of Natural language processing field, both research methods mentioned above can be used in designing the research. But the determinant is the subject of the research itself and the research question where both will determine the best-fitted method that the researcher can carry on. From my point of view, any productive research in natural language processing needs to build a model and do an experiment to check the model before start using it. So doing an experiment is essential. But how the experiment will be done, and what is the gab the researcher needs to fill. And even after determining what the gab is, the researcher needs the variable that he would include. So to conclude, Research in Natural language processing needs both. The researcher needs to start with LR or built on a survey to build breadth knowledge and to have an overview of all previous researches. Then the researcher needs to narrow his research and know specifically the detailed information about the model used, the variable, etc. After that, the researcher can follow it by experimental research as clear in the below framework in Fig. 2.

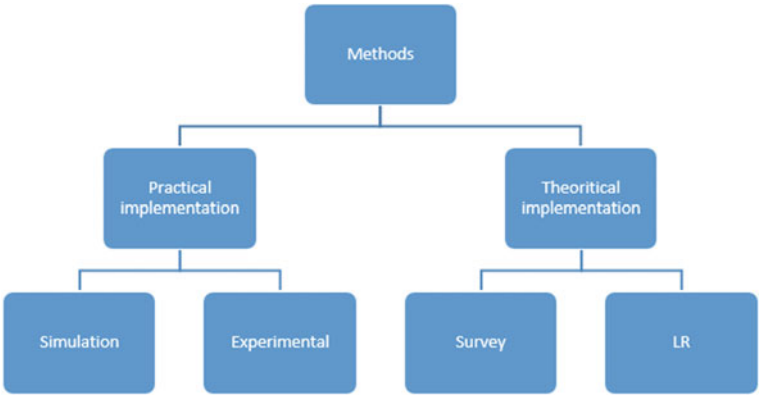


Fig. 1 Qualitative research design methods

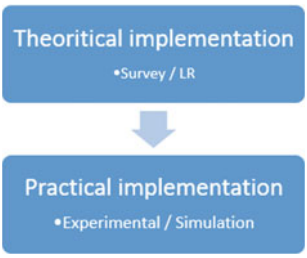


Fig. 2 Suggested framework for qualitative study in NLP

5 Conclusion

This study examined various peer-reviewed papers in the Natural Language Processing (NLP) field, aiming to illustrate the best methods used in text classification. Two research design methods illustrated in this study; survey and experimental research. The survey in the NLP field is not typical, it is more likely to systematic review for previous researches, and the researcher will determine his survey questions and start looking for their answers in the chosen paper. Whereat the end, it is the researcher's opinion. The advantages and disadvantages of both methods illustrated as well. To conclude, and from my point of view, doing research in NLP cannot be done without both of the mentioned research designs. The researcher needs to start with LR or built on a survey to build breadth knowledge and to have an overview for all previous researches to know the most popular methods, state of the art, and the gap. Then the researcher needs to narrow his research and know specifically the detailed information about the model used, the variable, etc. After that, the researcher can follow it through experimental research.

Acknowledgements This work is a part of a project submitted in fulfilment of Ph.D. Computer Science—Faculty of Engineering and Information Technology at The British University in Dubai.

References

1. Al Mansoori S, Almansoori A, Alshamsi M, Salloum SA, Shaalan K. Suspicious activity detection of Twitter and Facebook using sentimental analysis
2. Alsharhan A, Salloum S, Shaalan K. The impact of e learning as a knowledge management tool in organizational performance
3. Salloum SA (2021) Sentiment analysis in dialectal Arabic: a systematic review. In: Advanced machine learning technologies and applications: proceedings of AMLTA 2021, p 407
4. Salloum SA (2021) Classifying audio music genres using a multilayer sequential model. In: Advanced machine learning technologies and applications: proceedings of AMLTA 2021, p 301
5. Salloum SA (2021) Predicting the intention to use social media sites: a hybrid SEM-machine learning approach. In: Advanced machine learning technologies and applications: proceedings of AMLTA 2021, p 324

6. Almatrooshi F, Akour I, Alhammadi S, Shaalan K, Salloum SA. A recommendation system for diabetes detection and treatment
7. Yousuf H, Al-Hamad AQ, Salloum S. An overview on CryptDb and Word2vec approaches
8. Khamees AA, Hejazi HD, Alshurideh M, Salloum SA (2021) Classifying audio music genres using CNN and RNN. In: Advanced machine learning technologies and applications: proceedings of AMLTA 2021, p 315
9. Elnagar A, Yagi SM, Nassif AB, Shahin I, Salloum SA (2021) Systematic literature review of dialectal Arabic: identification and detection. IEEE Access 9:31010–31042
10. Alshamsi A, Bayari R, Salloum S. Sentiment analysis in English texts
11. Shaalan K, AlShuweihi M, Salloum SA (2021) Biomedical corpora and natural language processing on clinical text in languages other than English: a systematic review. In: Al-Emran M, Shaalan K, Hassanien A (eds) Recent advances in intelligent systems and smart applications, vol 295. Springer, Cham
12. Wahdan KSA, Hantoobi S, Salloum SA, Shaalan K (2020) A systematic review of text classification research based on deep learning models in Arabic language. Int J Electr Comput Eng 10(6):6629–6643
13. Salloum SA, Al-Emran M, Abdallah S, Shaalan K (2017) Analyzing the Arab Gulf newspapers using text mining techniques. In: International conference on advanced intelligent systems and informatics, pp 396–405
14. Salloum SA, Al-Emran M, Shaalan K (2016) A survey of lexical functional grammar in the Arabic context. Int J Com Net Tech 4(3)
15. Salloum SA, AlHamad AQ, Al-Emran M, Shaalan K (2018) A survey of Arabic text mining, vol 740
16. Etaiwi W, Awajan A (2018) Graph-based Arabic NLP techniques: a survey. Procedia Comput Sci 142:328–333
17. Elnagar A, Al-Debsi R, Einea O (2020) Arabic text classification using deep learning models. Inf Process Manag 57(1):102121
18. Bahassine S, Madani A, Kissi M (2016) An improved Chi-square feature selection for Arabic text classification using decision tree. In: SITA 2016—11th international conference on intelligent systems: theories and applications
19. Salloum SA, AlHamad AQ, Al-Emran M, Shaalan K (2018) A survey of Arabic text mining. In: Studies in computational intelligence, vol 740, pp 417–431. Springer, Cham
20. Hmeidi I, Al-Ayyoub M, Abdullah NA (2015) Automatic Arabic text categorization: a comprehensive comparative study. J Inf Sci 41(1):114–124
21. Patton DU, McKeown K, Rambow O, Macbeth J (2016) Using natural language processing and qualitative analysis to intervene in gang violence: a collaboration between social work researchers and data scientists. August
22. Safdar N, Abbo LM, Knobloch MJ, Seo SK (2016) Research methods in healthcare epidemiology: survey and qualitative research. Infect Control Hosp Epidemiol 37(11):1272–1277
23. (2016). 7 Advantages and disadvantages of experimental research. Flow Psychology

Heterogeneous Neural Collaborative Filtering for a Business Recommender System

Vu Hong Quan and Phan Duy Hung

Abstract In recent years, lots of techniques applied and optimized to make the recommender system better. Most of them mainly focus on the target interaction between users and items or a part of auxiliary information but hardly leverage other useful information relevant to customer transactions. In this paper, we want to propose a new model named Heterogeneous Neural Collaborative Filtering (HNCF) for learning recommender systems from two important parts: multi-aspects and multi-behaviors. The HNCF algorithm proposed is divided into four parts: Commuting similarity matrix, Multi-Layer Perceptron, Fusion by the Attention Mechanism, Multi-behavior Prediction. The model exploits characteristics of customers and properties of products from different aspects besides the aspect of purchase by building meta paths then commuting similarity matrices. Aspect-level latent factors fusion gives results as factors representing each user and item. These factors synthesizing with multi-behaviors prediction layers is the highlight of the model to make a better recommender system.

Keywords Multi-aspects · Multi-behaviors · Recommender system · Heterogeneous network

1 Introduction

1.1 Problem and Motivations

Current Nowadays, with the development of the global Internet network, customers of information systems, especially e-commerce websites, gradually get acquainted and regularly use electronic transactions in many fields such as securities, insurance,

V. H. Quan · P. D. Hung (✉)
FPT University, Hanoi, Vietnam
e-mail: hungpd2@fe.edu.vn

V. H. Quan
e-mail: quanvhhe130299@fpt.edu.vn

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_9

finance, banking, technology, telecommunications, and so on [1, 2]. However, with a huge amount of information, they often feel difficult to find and choose suitable information. For example, customers want to deposit money in the bank but hardly know which bank products to choose, whether they should choose an insured deposit product or not? What are the benefits and possible risks and what is the best product for the customer? Customers have lots of choices but not enough time or knowledge to make optimal decisions.

1.2 Related Works

The aspects of a product play an important role in the recommendation domain (e.g., the director and actors in a movie or the category of a product). Most existing latent factor models only utilize the rating information between users and items but are not concerned about other useful aspects [1], or employs ineffective methods to compute the similarities between aspects [2].

Existing approaches for multi-behavioral recommendation can be divided into two categories [3]. The first is the Collective Matrix Factorization (CMF) [4, 5], which jointly factorizes multiple behavior matrices based on extending the matrix factorization (MF) method. Another category approaches the problem from the perspective of learning [6]. It is natural to assume that items a user has interacted with should be more preferable when recommender models learn from the interaction data. However, these data often lack behavior semantics such as the add-to-cart behavior cannot happen before a view or a view can happen without purchase. Chen et al. [3] leveraged these factors and correlated the model prediction of each behavior type in a cascaded way but fixed the relational line in behaviors.

1.3 Contribution

This work proposes a new model named Heterogeneous Neural Collaborative Filtering (HNCF) which combines the two important factors: multi-aspect information and the translational relationships between different behaviors.

From a defined Heterogeneous Information Network (HIN), the model first determines the meta paths based on the structure of the dataset, then uses PathSim [7] to compute the similarity matrices between users with other users and items with other items. Aspect-level latent factors are learned through a standard Multi-Layer Perceptron (MLP) then fused by the attention mechanism. This paper focuses on aggregating these result factors with multi-behaviors prediction layers to improve the performance.

The data for training and testing was obtained from an online shop in the Middle East at the end of 2019. The efficient optimization method was implemented with TensorFlow and optimized using mini-batch Adagrad [8]. The proposed

algorithmic process is tested to verify their performance and effectiveness is confirmed using two indexes: Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG).

2 Methodology

2.1 Problem Formulation

Formally:

Let U be a set of n users, |U| = n, and u be a specific user ($u \in U$).

Let I be a set of m items, |I| = m, and i be a specific item ($i \in I$).

Let $\{y_1, y_2, \ldots, y_K\}$ denote the user-item interaction matrices for all the K types of behaviors, where $y_k = [y_{k-uv}] \in \{0, 1\}(u \in U$ and $i \in I)$ indicates whether user u has interacted with item v under behavior k. Generally, a model has a target behavior to be optimized, which is denoted as y_K. In our model, the target behavior is purchase and other behaviors are click, view, and add-to-cart.

2.2 Model Overview

First of all, meta paths containing objects are created based on the rich user-item interaction information in the recommendation, and PathSim is used to compute similarity matrices for each meta path of HIN. Then these matrices are used as inputs for a neural network to learn the aspect-level latent factors separately. They will be combined with the attention mechanism and each user or item will have its final latent factor. Simultaneously, we apply the transfer-based multi-behavior prediction to generate the prediction layer for the *K*th behavior (the target behavior) from other behaviors. Therefore, the likelihood that user u (p_u) will purchase item v (q_v) under *k*th behavior is estimated with:

$$\hat{y}_{K-uv} = h_K^T \times (p_u \odot q_v) \tag{1}$$

2.3 Heterogeneous Information Network

The recommendation domain comprises various object types (e.g., author, paper, venue, and title in research paper recommendation) with rich relations among them, which establish a natural HIN. Therefore, a lot of techniques have been developed to leverage its heterogeneity [9]. An important role in HIN, the meta path [10], the

association of two items using a succession of relations between object types, is broadly used to create rich semantic data.

For example, a path that goes from Author → Paper → Venue → Paper → Author, where the relationships are 'write', 'public', 'public', 'written' respectively, is considered as a form of meta path.

In fact, there are several methods to measure the similarity of objects based on paths in a heterogeneous information network. ObjectRank [2] recognizes that heterogeneous relationships could have an impact on random walking, and assigned different propagation factors to each form of object relationship to derive a revised version of P-PageRank (ObjectRank). Personalized PageRank [11] is an asymmetrical measure of similarity which evaluates the likelihood of encountering object Y from object X by walking randomly on the network with a restart. Nevertheless, such methods only grant one specific combination of all the possible meta paths utilizing the fixed weights decided by the damping factor and propagation factors between different types.

Motivated by Sun et al. [7], this model uses PathSim similarity which finds peer objects in the network to measure the similarity between users versus users and items versus items based on their meta paths.

2.4 Processing Latent Factors

After calculating the feature vector of each user that denotes the similarity of this user to other users underneath each aspect (similar to each item), we need to project each of these vectors down a new space with the basic assumption is that there exist an unknown low-dimensional representation of users and items where the user-item relationship can be modeled accurately.

Multi-layer Perceptron

To have a network that is able to learn different aspect-level latent factors separately, this model employs a Multi-layer Perceptron (MLP) [12] which gives it a large level of flexibility and non-linearity. In addition to the Input and Output layers, an MLP can have multiple hidden layers in between.

Attention Mechanism

After applying MLP to learn the aspect-level latent factors separately for users and items within their domain, we need to incorporate them together to acquire upholstery factors. The attention-based aspect-level latent factors fusion is chosen for this purpose due to its proven effectiveness in various domains such as machine translation and image captioning [13, 14].

In general, there are two kinds of attention mechanisms: hard attention (reinforcement learning to learn where to pay attention) and soft attention (weighted learning using the backpropagation algorithm). The soft attention is chosen for our model due to the ease of optimization. The mechanism will learn the weights to pay

attention to on all the informative parts of the image, the sentence, or whatever is necessary to synthesize all the information to make a prediction. Specifically, given the user's category-aspect latent factor u_i^C, we use a two-layer network to compute the attention score s_i^C by the following equation:

$$s_i^C = \mathrm{W}_2^T \times f(\mathrm{W}_1^T \times u_i^C + b_i) + b_2 \tag{2}$$

where W_* and b_* are the weight matrix and bias, respectively.

Finally, we can apply the Softmax function and a normal ratio to obtain the attention value for the aspect-level latent factors:

$$w_i^C = \frac{\exp(s_i^C)}{\sum_{k=1}^{A} \exp(s_i^k)} \tag{3}$$

where A is the number of aspects.

With each k-aspect latent factor of user u_i, we have attention weights w_i^k respectively, the aggregated latent factor u_i can be calculated by the following equation:

$$u_i = \sum_{k=1}^{A} \mathrm{w}_i^k \times u_i^k \tag{4}$$

2.5 Multi-behavior Prediction

To demonstrate the importance of other behaviors of users besides target behavior is the purchase, we denote h_k as the prediction layer for the k_{th} behavior. As mentioned above, behaviors of users as view, add-to-cart, remove-from-cart, or purchase do not exist independently but are closely related. The prediction of a behavior is dependent on the predictions of its precedent behaviors [15]. Simultaneously, a transfer mechanism is also designed in knowledge representation [16]. Motivated by that mechanism, the transfer scheme of two relational behaviors (from h_t to h_k) is formulated as:

$$f_{h_t \to h_h} = h_t \times M_{tk} + r_{tk} \tag{5}$$

where M_{tk} is a transfer matrix projecting h_t from the t_{th} behavior space to the k_{th} behavior space and r_{tk} is the bias.

When we aggregate all of the behaviors on a sequence of behaviors to some considered behavior, we obtain the equation:

$$h_k = \sum_t (h_t \times M_{tk} + r_{tk}) \tag{6}$$

where t is considered as the precedent behaviors of that k_{tk} behavior. Of course, the first behavior (view) is randomly initialized.

Because of the sparsity of most E-commerce datasets, this prediction layer can make the recommender system better for users who have few behaviors and transactions.

2.6 Objective Function

The goal is to predict whether the user u_i interacts with item v_j or not under the kth behavior h_k in the range of [0, 1], especially the target behavior $\hat{y}_{K-ij}$ (purchase).

The probability of the interaction between the user u_i and item v_j is formulated according to the equation:

$$y_{k-ij} = sigmoid(h_k \times (u_i \odot v_j)) \tag{7}$$

Further assuming that the data points are randomly generated independently of each other, we could write:

$$p(\alpha, \alpha^- | w) = \prod_{i,j \in \alpha} \hat{y}_{ij} \prod_{i,k \in \alpha^-} (1 - \hat{y}_{ik}) \tag{8}$$

$$p(\alpha, \alpha^- | w) = \prod_{i,j \in \alpha \cup \alpha^-} (\hat{y}_{ij})^{y_{ij}} \prod_{i,k \in \alpha^-} (1 - \hat{y}_{ik})^{1 - y_{ij}} \tag{9}$$

where w is the parameter, α and α^- are the positive and negative instance sets, respectively.

Loss function respectively to k_{tk} behavior:

$$L_k = \sum_{i,j \in \alpha \cup \alpha^-} (y_{k-ij} \times \log \hat{y}_{k-ij}) + (1 - y_{k-ij}) \times \log(1 - \hat{y}_{k-ij})) \tag{10}$$

Loss function should be fused by a sum weighted to maintain the different level of behaviors:

$$L = \sum_{k=1}^{K} \lambda_k L_k \tag{11}$$

where $\sum_{k=1}^{K} \lambda_k = 1$.

3 Implementation and Analysis

3.1 Experimental Settings

Dataset

This model leverages a real-world E-commerce dataset from a large multi-category online store in the Middle East. The data was collected from October to November 2019. Auxiliary information is the category and brand of items. There are also 3 kinds of user behaviors: view, add-to-cart, and purchase. The original data is large but highly sparse (Table 1).

Evaluation

The model will employ the leave-one-out evaluation which was introduced former [17] to estimate how the model works for the recommender system. Specifically, it keeps the final interaction of the user for testing and the remaining for training. For each user, 99 unrelated items are used as negative samples. Then each user will have 1 set of 100 items for testing (100 couples of users and items). We use both the Hit Ratio (HR) and Normalized Discounted Cumulative Gain (NDCG) to evaluate the model performance. They are defined as:

$$HR = \frac{|\text{keys}|}{|\text{users}|} \tag{12}$$

$$NDCG = \frac{1}{|users|} \sum_{i=1}^{|users|} \frac{\log(p_i + 1)}{\log 2} \tag{13}$$

Table 1 Statistics of evaluation dataset

Interaction	Items	Users	Sparsity
82,533	6676	10,000	99.88%

where |keys| is the number of users whose test item appears in the recommended list and p_i is the position of the test item in the list for the ith test user.

In this evaluation, the model will proceed with HR@15 and NDCG@15. It means that HR and NDCG measure whether the test item is presented on the top-15 list.

3.2 Performance Comparison

Comparing HNCF and NeuCF

The NeuCF model was implemented [18] to test this dataset. Both HR@15 and NDCG@15 show that HNCF outperformed NeuCF. The NeuCF model still performs on the movieLen-1M dataset, but with our dataset, it's nearly inefficient. The most significant reason for this is due to the sparsity of this online shop (99.88%) being higher than the sparsity of MovieLen-1M (95.53%). The HNCF model is still good because it additionally incorporates lots of user behaviors that addressed the sparsity of the dataset (Figs. 1 and 2).

Analysis of the Attention Mechanism

When fusing latent factors, two methods are compared:

- Average: Averaging all the aspect-level latent factors
- Attention mechanism: weighted averaging of all pieces of information (Figs. 3 and 4).

The two charts show that the attention mechanism operates more effectively than average. This is due to the average method normalizing all of the important levels while the attention mechanism concentrates on what is significant.

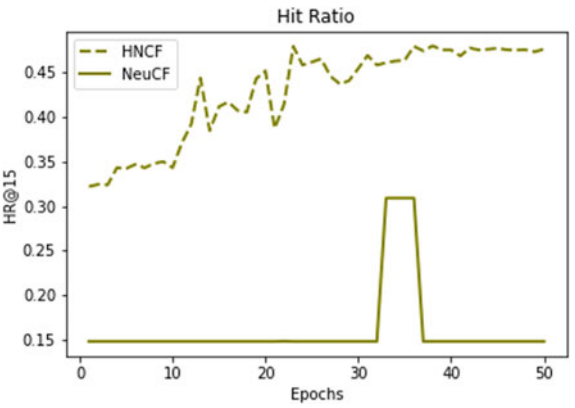


Fig. 1 HR of 2 models

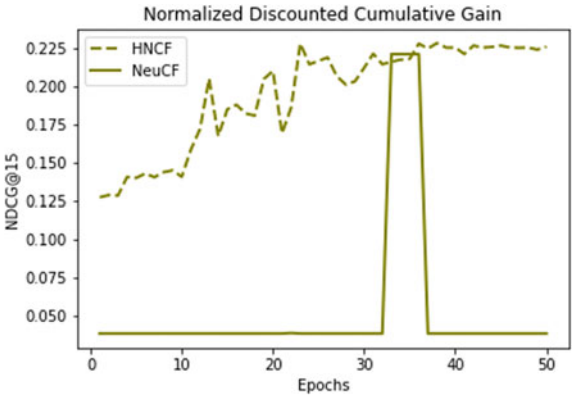


Fig. 2 NDCG of 2 models

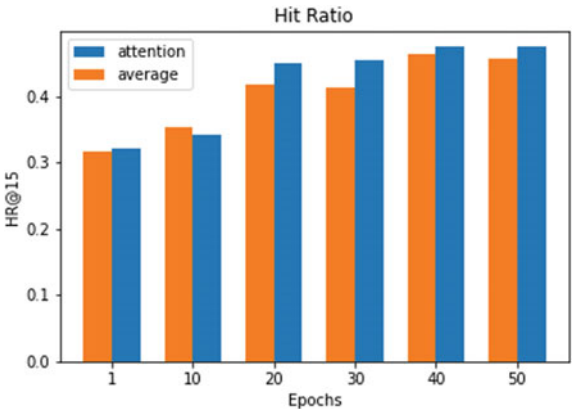


Fig. 3 HR when using attention mechanism and average method

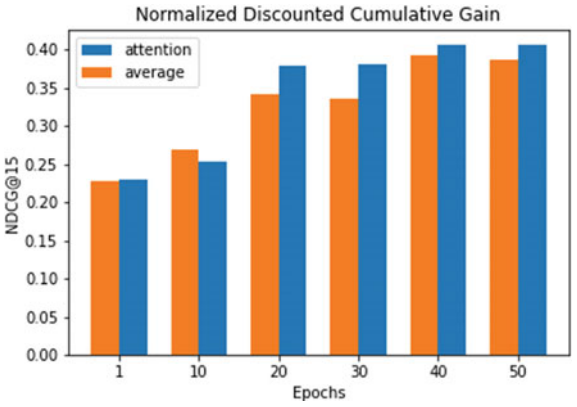


Fig. 4 NDCG when using attention mechanism and average method

Adjusting the Number of Layers

The figure demonstrates that the more the number of layers is used (2, 3 and 4 layers), the more HR@15 and NDCG@15 also increased. However, when the model increases from 3 to 4 layers, the performance of the model is just a little bit better. Analogously, when MLP is implemented with more than 4 layers, both HR@15 and NDCG@15 increased trivially. Therefore, for a standard Multi-layer Perceptron (MLP), we should use 4 layers to achieve the algorithm's effectiveness (Figs. 5 and 6).

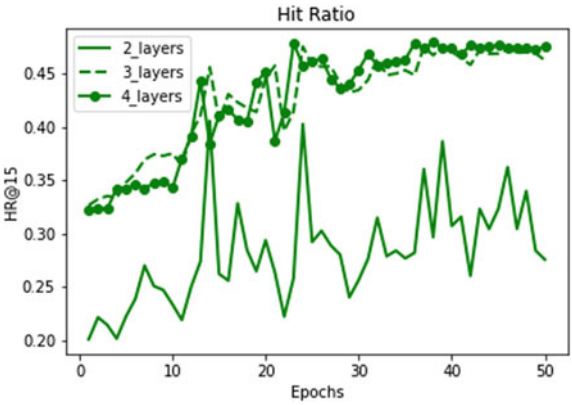


Fig. 5 HR when MLP uses 2, 3, or 4 layers

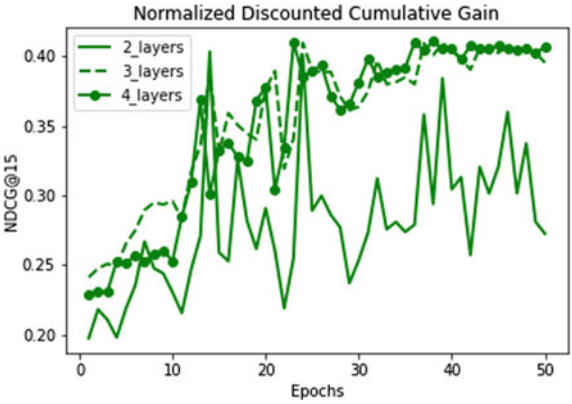


Fig. 6 NDCG when MLP uses 2, 3, or 4 layers

4 Conclusion and Perspectives

To increase the efficiency of e-commerce sites and online shops, Ecommerce sites integrate the recommender system into the main system to be able to suggest to users the most suitable choices for them, leading to increasing the number of successful transactions. Processing a huge amount of data as well as applying machine learning and deep learning techniques play an important role in making a better system.

Most of the models focus on explicit interaction between users and items and forget other useful information. The others either focus on different aspects, e.g., the category-aspect, brand-category of items, or only handle heterogeneous feedback (i.e., view, click, add to cart) without leveraging all of the useful properties of the amount of that data.

This paper proposes a new model named HNCF that incorporates processing techniques to increase the efficiency of the recommender system. The model first calculates the similarities of users to each other and items through aspect-level latent factors. After aggregating the results, the factors are processed together with the multi-behavior prediction layers to give a probability of the interaction between the user U_i and the item V_j is the highlight of the model to enhance the performance of the recommender system. Each user is given then 100 items to test the performance of the algorithm.

This study can be applied to most e-commerce sites because it solves the basic problems of online shopping such as Cold-start and sparsity [19, 20]. Multi-behaviors property enables it to be applied in the Movie recommendation domain. Besides, techniques of exploiting lots of aspects can be helpful in social networks to suggest friends and content to users [21, 22].

References

1. Gosh S, Nahar N, Wahab MA, Biswas M, Hossain MS, Andersson K (2020) Recommendation system for e-commerce using alternating least squares (ALS) on apache spark. In: Proceedings of the 3rd international conference on intelligent computing and optimization
2. Balmin A, Hristidis V, Papakonstantinou Y (2004) ObjectRank: authority-based keyword search in databases. In: VLDB, pp 564–575
3. Gao C, He X, Gan D, Chen X, Feng F, Li Y, Chua T-S, Yao L, Song Y, Jin D (2019) Learning to recommend with multiple cascading behaviors. In: IEEE transactions on knowledge and data engineering
4. Park C, Kim D, Oh J, Yu H (2017) Do "also-viewed" products help user rating prediction? In: Proceedings of the 26th international conference on world wide web. International world wide web conferences steering committee, pp 1113–1122. Republic and Canton of Geneva, CHE
5. Zhao Z, Cheng Z, Hong L, Chi EH (2015) Improving user topic interest profiles by behavior factorization. In: Proceedings of the 24th international conference on world wide web. International world wide web conferences steering committee, pp 1406–1416. Republic and Canton of Geneva, CHE

6. Bauer J, Jannach D (2018) Optimal pricing in e-commerce based on sparse and noisy data. In: Decision support systems, vol 106, pp 53–63. ISSN 0167-9236
7. Sun Y, Han J, Yan X, Yu PS, Wu T (2011) PathSim: meta path-based top-K similarity search in heterogeneous information networks. Proc VLDB Endow 4:992–1003
8. Ward R, Wu X, Bottou L (2019) AdaGrad stepsizes: sharp convergence over nonconvex landscapes. In: Proceedings of the 36th international conference on machine learning, PMLR 97, pp 6677–6686
9. Shi C, Zhang Z, Luo P, Yu PS, Yue Y, Wu B (2015) Semantic path based personalized recommendation on weighted heterogeneous information networks. In: Proceedings of the 24th ACM international on conference on information and knowledge management (CIKM' 15). Association for computing machinery, pp 453–462. New York, NY, USA
10. Liu X, Yu Y, Guo C, Sun Y (2014) Meta-path-based ranking with pseudo relevance feedback on heterogeneous graph for citation recommendation. In: Proceedings of the 23rd ACM international conference on conference on information and knowledge management (CIKM' 14). Association for computing machinery, pp 121–130. New York, NY, USA
11. Jeh G, Widom J (2003) Scaling personalized web search. In: Proceedings of the 12th international conference on World Wide Web. Association for computing machinery, pp 271–279. New York, NY, USA
12. Mirjalili S (2019) Evolutionary multi-layer perceptron. In: Evolutionary algorithms and neural networks. Studies in computational intelligence, vol 780. Springer, Cham. https://doi.org/10.1007/978-3-319-93025-1_7
13. Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y (2015) Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning, pp 2048–2057
14. You Q, Jin H, Wang Z, Fang C, Luo J (2016) Image captioning with semantic attention. In: IEEE conference on computer vision and pattern recognition, pp 4651–4659
15. Chen C, Zhang M, Zhang Y, Ma W, Liu Y, Ma S (2020) Efficient heterogeneous collaborative filtering without negative sampling for recommendation. AAAI 34(01):19–26
16. Bordes A, Usunier N, Garcia-Durán A, Weston J, Yakhnenko O (2013) Translating embeddings for modeling multi-relational data. In: Proceedings of the 26th international conference on neural information processing systems, vol 2, pp 2787–2795. Curran Associates Inc., Red Hook, NY, USA
17. Bayer I, He X, Kanagal B, Rendle S (2017) A generic coordinate descent framework for learning from implicit feedback. In: Proceedings of the 26th international conference on world wide web. International world wide web conferences steering committee, pp 1341–1350. Republic and Canton of Geneva, CHE
18. Shi C et al (2019) Deep collaborative filtering with multi-aspect information in heterogeneous networks. In: IEEE transactions on knowledge and data engineering
19. Hung PD, Huynh LD (2019) E-commerce recommendation system using mahout. In: Proceedings of the IEEE 4th international conference on computer and communication systems (ICCCS), Singapore, pp 86–90. https://doi.org/10.1109/CCOMS.2019.8821663
20. Hung PD, Ngoc ND, Hanh TD (2019) K-means clustering using R: a case study of market segmentation. In: Proceedings of the 5th international conference on e-business and applications (ICEBA 2019). Association for computing machinery, New York, NY, USA, pp 100–104. https://doi.org/10.1145/3317614.3317626
21. Hung PD, Lien NTT, Ngoc ND (2019) Customer segmentation using hierarchical agglomerative clustering. In: Proceedings of the 2nd international conference on information science and systems (ICISS 2019). Association for computing machinery, New York, NY, USA, pp 33–37. https://doi.org/10.1145/3322645.332267
22. Hung PD, Dat DQ (2020) Analysing effects of customer clustering for customer's account balance forecasting. In: Nguyen NT, Hoang BH, Huynh CP, Hwang D, Trawiński B, Vossen G (eds) Computational collective intelligence, ICCCI 2020. Lecture notes in computer science, vol 12496. Springer, Cham. https://doi.org/10.1007/978-3-030-63007-2_20

An Online Clustering Approach for Evolving Data-Stream Based on Data Point Density

Md Kamrul Islam and Bishnu Sarker

Abstract Density-based clustering has been widely studied by many researchers due to its popularity in generating arbitrary shaped clusters in data-stream. However, the majority of them are offline, sliding windowed or online-offline hybridization and cannot be applied to data-stream applications. This paper presents a fully online density-based clustering approach namely OCED (Online Clustering for Evolving Data-stream). The algorithm summarizes the data points in a sphere-shaped data structure called a micro-cluster. The initial radius of micro-cluster is learned and updated towards its optimal value based on the average data point density. A cluster graph is generated in an online manner using the micro-clusters to provide the arbitrary shaped cluster. The micro-clusters, which are irrelevant to recent data-stream content, are identified and removed to handle the evolving property. To the best of our knowledge, this is the first clustering approach that learns the radius instead of predefining. The approach is evaluated in a wide range of experiments. OCED proves its superiority over aligned approaches in terms of computational speed, cluster quality, and memory efficiency.

Keywords Data-stream · Density-based clustering · Micro-cluster

1 Introduction

The unbounded series of data points is referred to as data-stream, which arrive continuously from an application like Facebook, Amazon, automated teller machine (ATM) [1]. In the field of data-stream mining, clustering is an essential tool for finding frequent patterns (clusters). Based on how data-streams are processed, clustering approaches are categorized into online and offline categories [2]. The requirement of

M. K. Islam (✉) · B. Sarker
Universite de Lorraine, Nancy, France
e-mail: kamrul.islam@loria.fr

M. K. Islam
Jashore University of Science and Technology, Jashore, Bangladesh

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_10

storing the whole data-stream in memory makes an offline clustering unrealistic to apply to data-stream applications as data-stream has uncertainty in its velocity, volume, and variety [3]. For handling these three uncertainties, online clustering approaches fetch data points sequentially, find the suitable cluster for them, and remove the data points immediately from memory. An ideal online clustering approach for data-stream tries to generate high quality and arbitrary shaped clusters in minimum processing time in both clean and noisy environments [4]. The approach should also evolve to handle the variety property of data-stream, and the recent snapshot of the approach should reflect the recent contents of the data-stream. Based on the working principle, the clustering approaches are further divided into five major categories, and they are partitioning, hierarchical, density-based, grid-based, and model-based approaches [4, 5]. In recent years, density-based approaches have gained much attention from researchers due to their success in generating high-quality and arbitrary shaped clusters in data-stream. DBSCAN [6] is considered to be the primitive density-based approach for generating arbitrary shaped clusters in a large database. DBSCAN is extended to DenStream [7] and CVD-Stream [8] for clustering of data-stream, which consist of online and offline phases. The online phase summarizes the data points in micro-clusters, and the clusters are generated in the offline phase. Based on DenStream, several clustering approaches have been developed, such as C-Denstream [9], VDstream [1] to generate arbitrary shaped clusters in data-stream. However, these approaches are not fully online approaches. ELM [10], DEC [11], and CODAS [12] are three fully on-line clustering approaches. However, ELM and DEC can not generate arbitrary shaped clusters, and CODAS can not capture the evolving property. The limitation of CODAS is overcome in CEDAS, which handles the evolving property for data-stream by introducing a cluster aging procedure. All of the approaches mentioned above use a user-defined and application-dependent micro-cluster radius. The global setting of the radius may not be optimal for all micro-clusters and thus degrades the cluster performance. This problem is reduced in BOCEDS, which uses the concept of maintaining a local optimal radius instead of a global radius of a micro-cluster [13]. However, the memory requirement is high in BOCEDS as it stores outliers for future reference, and the overlapping clusters are not properly managed. Moreover, the uncontrolled growth of micro-cluster radius causes some micro-clusters to grow in sparse regions. This paper presents an online approach of density-based clustering to overcome the difficulties mentioned above, namely OCED. The specific contributions of this article are OCED

- proposes a fully online approach for clustering of evolving data-stream by maintaining local optimal radius.
- learns the minimum radius of micro-cluster instead of predefining to reduce the dependency on the user in setting the radius and the chance of erroneous setting of the radius.
- reduces the formation of clusters in the sparse region to improve cluster quality by updating the radius of micro-cluster towards the local optimal based on data point density.
- introduces a micro-clusters merging procedure to improve memory efficiency.

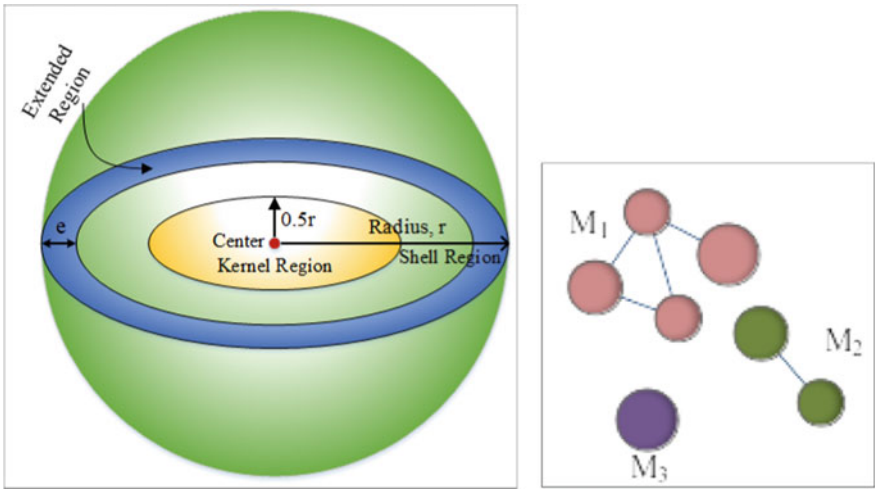
2 The Proposed OCED Approach

Micro-cluster is the heart of OCED, which summarizes the data points from a data-stream. A cluster graph is formed based on the intersection of micro-clusters in the data-space. The cluster graph is composed of one or more disjoint sub-graphs where the nodes represent the micro-clusters, and the edge represents the intersection of two micro-clusters. Each sub-graph is termed a macro-cluster or simply cluster.

2.1 Data Structures in OCED

The structure of a sphere-shaped micro-cluster is illustrated in Fig. 1a. Based on the intersection of micro-clusters, the generated cluster graph is illustrated in Fig. 1b. A micro-cluster (M) is simply a tuple MC(C, R, N, NE, E, L, M), where center(C) is the average of the data points, radius (R) is the spread from the center and local count (N) is the number of data points in the micro-cluster.

Each micro-cluster is divided into kernel and shell regions. The kernel region is the outer half region (covered by 0.5R), whereas the shell region is the remaining inner half region. When the kernel region of one micro-cluster intersects another micro-cluster in any region, they are said to be edged. The edge list (L) of a micro-cluster is the collection of such edged micro-clusters. Extended count (NE) is the amount of data points in the extended region, energy (E) is the potential, and



(a) Micro-cluster (b) Cluster graph

Fig. 1 Data structures in OCED. **a** The structure of a micro-cluster with the labeling of each part, **b** the formation of three macro-clusters from seven micro-cluster based on their connectivity where circles represent micro-clusters and a line represents the edge between micro-clusters

macro-cluster (M) is an identifier of the macro-cluster or cluster of the micro-cluster. OCED uses E to decide whether a micro-cluster should live or die. Figure 1b shows the formation of a cluster graph where each sub-graph is illustrated with distinguishable color. Each sub-graph in the cluster graph is referred to as macro-cluster or simply cluster. Figure 1b shows the generation of three macro-clusters (M_1, M_2, and M_3) from seven micro-clusters.

OCED defines three types of micro-clusters namely potential micro-cluster, core micro-cluster, and outlier micro-cluster based on a user-defined count threshold. A core micro-cluster is a micro-cluster in true data point space which has a local count above or equal to the threshold and positive energy whereas a potential micro-cluster has a local count below the threshold and positive energy. Outlier micro-clusters have zero or negative energy and they are either background noise or irrelevant to current data-stream content. They are removed immediately after identification. A potential micro-cluster converts to either core micro-cluster or outlier micro-cluster in later. Among the three types of micro-clusters, only core micro-clusters are used to generate macro-clusters.

2.2 *Description of the Algorithm*

OCED defines three application-dependent parameters before starting the execution of the approach. The first one is *Decay* which is defined as data point arrival rate per unit time. The second one, the count threshold (N_{th}), is defined as the least count of data points needed to generate a core micro-cluster. The final parameter is radius growth (R_{gr}) which defines how fast the radius of a micro-cluster grows towards its local optimal. Once the parameters are set, OCED starts fetching the data points and generates clusters in eight distinct procedures.

Initialization of algorithm. After defining the three application-dependent parameters based on expert knowledge, OCED starts with initializing a minimum radius value. The algorithm set the initial radius as the radius growth (R_{gr}), assuming that a micro-cluster has enough spread to get local count above the count threshold to be a core-micro-cluster. This parameter is learned after an incubation (or observation) period (described later). The volume, V of an S-dimensional micro-cluster with radius R_{min} is computed as Eq. (1). The minimum data-point density is the average data point density which is computed based on the volume of the micro-cluster and count threshold as Eq. (2).

$$V = \frac{\pi^{\frac{s}{2}}}{\Gamma\left(\frac{S}{2}+1\right)}(R_{\min})^S \tag{1}$$

$$D_{\min} = \frac{V}{N_{th}} \tag{2}$$

where Γ is a gamma function.

Target micro-cluster searching. When a data point X_{t+1}, arrives at a time (t + 1), OCED seeks the micro-cluster where the data point stays by comparing the Euclidean distance (d) from X_{t+1} to each alive micro-cluster. The data point X_{t+1} resides inside a micro-cluster with center (C) and radius (R) if the following condition is satisfied.

$$d(X_{t+1}, C) \leq R \tag{3}$$

The mapped micro-cluster is either a potential micro-cluster or a core micro-cluster. Potential micro-clusters have the priority over core micro-cluster to en-sure generating more core micro-cluster. If X_{t+1} does not reside in any of the micro-cluster regions in the current model, it creates a new potential micro-cluster itself.

Micro-cluster creation. A data point X_{t+1}, which falls into the empty region in the data space, creates a potential micro-cluster itself by setting the data point itself as the center ($C = X_{t+1}$), local count to 1(i.e. N = 1), extended count to 0(i.e. NE = 0), radius to minimum radius ($R = R_{min}$), empty edge list (L = Ø), energy to 1 (i.e. E = 1.0), and default macro-cluster id (i.e. M = 0).

Micro-clusters updating. When a micro-cluster MC(Ct, Rt, Nt, Net, Et, Lt, Mt) gets a data point X_{t+1}, its data structure is updated in an online manner. The local count, center, and energy are updated similarly to the BOCEDS approach [13].

$$N_{t+1} = N_t + 1 \tag{4}$$

$$E_{t+1} = E_t + \frac{\{R_t - d(X_{t+1}, C)\}}{R_t} \times \frac{1}{Decay} \tag{5}$$

$$C_{t+1}^k = \frac{\{N_{t+1} - 1\} \times C_t^k + X_{t+1}^k}{N_{t+1}} \tag{6}$$

for k = 1, 2, 3, …, S where S is the data point dimension size. If MC is a core micro-cluster and X_{t+1} falls in the extended region of MC, then its extended count is incremented by 1 (i.e. $NE_{t+1} = NE_{t+1}$). The volume (ΔV) and the density (D_{ex}) of the extended region of the micro-cluster are computed as

$$\Delta V = \frac{\pi^{\frac{S}{2}}}{\Gamma\left(\frac{S}{2} + 1\right)} \left\{ (R_t)^S - \left(R_t - R_{gr}\right)^S \right\} \tag{7}$$

$$D_{ex} = \frac{\Delta V}{NE_{t+1}} \tag{8}$$

In the case of $D_{ex} \geq D_{min}$, MC's radius is increased by R_{gr} as in Eq. 9, and the extended count is set to 0 (i.e. $NE_{t+1} = 0$). This controlled growth in radius prevents the micro-cluster from growing in spare regions.

$$R_{t+1} = R_t + R_{gr} \tag{9}$$

Otherwise, if MC is a potential micro-cluster then it is converted into a core micro-cluster if its local count meets the count threshold ($N_{t+1} \geq N_{th}$). Due to frequent shifting of micro-cluster center, a bigger micro-cluster, $MC_1(C_1, R_1, N_1, NE_1, E_1, L_1, M_1)$ may contain a smaller micro-cluster, $MC_2(C_2, R_2, N_2, NE_2, E_2, L_2, M_2)$ if the following condition (Eq. 10) is held.

$$R_1 = d(C_1, C_2) + R_2 \tag{10}$$

In case of holding the above merging condition (Eq. 10), the micro-clusters are merged by removing MC_2 and updating the data structure of MC_1 as Eqs. (11–14).

$$E_1 = \max(E_1, E_2) \tag{11}$$

$$C_1^k = \frac{N_1 \times C_1^k + N_2 \times C_2^k}{N_1 + N_2} \tag{12}$$

$$N_1 = N_1 + N_2 \tag{13}$$

$$NE_1 = NE_1 + NE_2 \tag{14}$$

Outlier identification. Each micro-cluster energy is reduced by an amount of (1/Decay) to identify outliers (noise/irrelevant) concerning recent data-stream contents. A micro-cluster with less or equal to zero energy is considered as an outlier and removed from the current model.

Cluster graph updating. OCED maintains the clustering graph based on the edging of core-micro-clusters to generate the macro-cluster or simply cluster in an online manner. The clustering graph needs to be updated if new core micro-clusters are generated, or core micro-clusters are removed, or the center of a core micro-cluster is shifted. For updating the cluster graph, the edge lists of micro-clusters are recomputed. Two core-micro-clusters, MC_1, MC_2 with radius R_1, R_2 and centers C_1, C_2, are edged in case of holding the following relation (Eq. 15).

$$d(C_1, C_2) \leq \max(R_1, R_2) + \frac{1}{2}\min(R_1, R_2) \tag{15}$$

where $d(C_1,C_2)$ is the Euclidean distance between micro-cluster centers MC_1 and MC_2. The macro-clusters are recomputed from the updated cluster graph.

Minimum radius learning. We assume that the initial minimum radius parameter is less than optimal, and it is updated towards the optimal in a period called the incubation period. At the end of the incubation period, OCED checks whether at least one core micro-cluster exists in the current model. If such a micro-cluster

exists, then the initial minimum radius serves as the mini-mum radius for the rest periods. Otherwise, OCED approach learns it through a learning process. The learning process starts with increasing the radii of micro-clusters by an amount of R_{gr}. In the next step, the overlapped micro-clusters are merged using Eqs. (11–14). Two overlapping micro-clusters intersect each other in their kernel regions. The process is repeated until at least one core micro-cluster is formed. The radius of the new core micro-cluster is set as the minimum radius (R_{min}). This minimum radius is used for the rest periods for micro-cluster initiation.

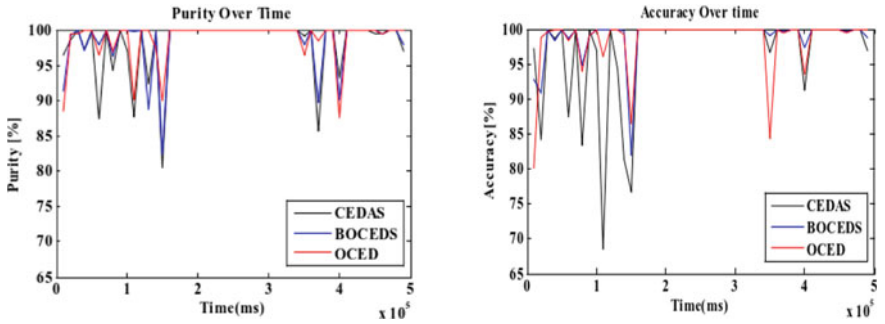
3 The Proposed OCED Approach

MATLAB R2014a is used to implement the proposed OCED approach. The approach is run in an environment with core i7 processor and 8 GB memory. The source code will be available on Github (https://github.com/kamrul-loria/OCED). The performance metrics include cluster quality, computational time, and memory efficiency. The performance of OCED is compared with two recently developed fully online density-based approaches, namely BOCEDS and CEDAS.

3.1 Cluster Quality

A quality metric of a cluster is described with the accuracy and purity of the cluster [13, 14]. The number of data points in a dominant cluster is referred to the cluster purity. However, the purity metric alone can not describe the cluster quality as an approach can show high purity value though a majority of data points are not clustered correctly. Cluster accuracy is the amount of truly clustered data points for overcoming the limitation of purity. KDDCUP'99 is a popular benchmark data-stream for measuring cluster quality [12, 15]. The data-stream contains 490,000 network traffics representing intrusion attacks. Each of the traffic consists of 41 features with a label of traffic. For clustering of KDDCUP'99 data-stream, OCED is configured with decay to 1000, the count threshold to 3 as in BOCEDS. The radius growth rate is set to 0.01. The purity and accuracy of OCED are measured at an interval of 10 K data points. Besides, the purity and accuracy of CEDAS and BOCEDS are taken from the results presented in [13]. Figure 2a, b compare the purity and accuracy respectively of OCED with CEDAS and BOCEDS. Figure 2a, b show that cluster purity and accuracy of OCED are lower than BOCEDS and CEDAS in an initial incubation period.

The minimum radius is learned using a radius learning procedure in OCED in this period. The cluster purity and accuracy are improved after this learning period. The accuracy and purity values of the CEDAS approach are lowest in almost all of the period as CEDAS uses a global micro-cluster radius which may grow in sparse regions. On the other hand, BOCEDS and OCED use the local radius concept,



(a) Cluster purity (b) Cluster accuracy

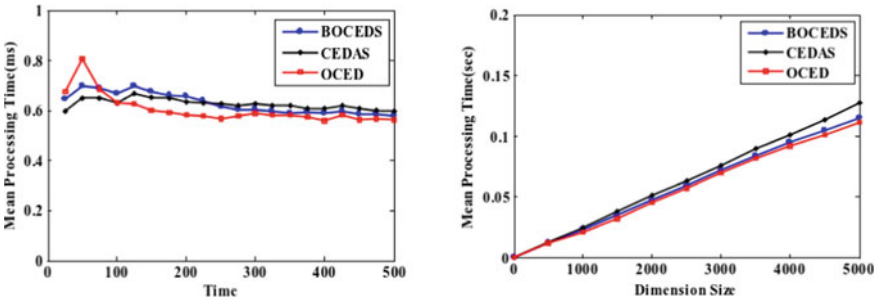
Fig. 2 Purity and accuracy for clustering of KDDCUP'99 data-stream. **a** The overall purity for OCED, BOCEDS, and CEDAS are 0.96, 0.96, and 0.94 respectively **b** the overall accuracy of OCED, BOCEDS, and CEDAS are 0.96, 0.95, and 0.95 respectively

which is updated towards its optimal as a new data point comes in the cluster. However, the uncontrolled growth process of the radius in BOCEDS provokes some micro-clusters to grow in sparse regions. This limitation is overcome in OCED by growing the micro-cluster radius for a minimum data point density. Thus, OCED generates more pure and accurate clusters than BOCEDS after the incubation period. In Fig. 2a, b, the purity and accuracy values in OCED remain highest in nearly all of the rest periods among all approaches. In Fig. 2, the purity and accuracy of OCED remain higher than 90% at all the period except the period 150 and 370. The frequent change in contents of data-stream causes these falls in quality performance. However, the purity values of OCED are still higher than BOCEDS and CEDAS in these two periods. Thus, OCED provides highly pure and accurate clusters.

3.2 Processing Time

Describing the processing speed of OCED, two popular sets of helical data-stream are used [13, 14, 16]. The first one is a time series of 5000 data points representing a set of three-dimensional circular helix [17]. This data-stream is used to measure the mean processing time to complete the clustering of data-stream. For evaluating the scalability of OCED, higher dimensional data-streams are generated by moving the data points into higher dimensional data space, as explained in [13]. In both experiments, decay and count threshold are set to 1000 and 4, as in [13]. The radius growth rate is set to 0.01. The experimental results are compared to two fully only clustering approaches, CEDAS and BOCEDS in Fig. 3.

From Fig. 3a, the initial processing time of OCED is the highest among the three approaches as the initial micro-cluster radius in OCED is less than the optimal



(a) 3D helical data stream (b) Highd imensional helical data stream

Fig. 3 Computational time for helical data-stream. **a** The change in mean processing time as new data points arrive from low dimensional helical data-stream with time **b** the growth of mean processing time as the dimension size of high dimensional helical data-stream increases

radius. Due to this fact, OCED generates more micro-clusters and takes more processing time in this period. As time progresses, the radii grow towards their optimal. The total amount of micro-clusters falls in OCED as well as in BOCEDS. However, the falling trend is more in OCED than BOCEDS as OCED introduces a procedure for merging micro-clusters for reducing the number of micro-clusters. On the other hand, CEDAS shows a stable processing time in the rest of the periods. The processing time of OCED remains the lowest among the three approaches for the rest of the periods. Figure 3b shows the scalability of clustering approaches when applied to high dimensional data-stream. This scalability experiment shows which the mean processing time linearly increases with increasing dimension size in all of the mentioned approaches. Thus, the processing time of OCED is lower than CEDAS and BOCEDS due to the same reason as described above, which supports the superiority of OCED over CEDAS and BOCEDS when considering the scalability metric.

3.3 Memory Efficiency

An online clustering approach defines its memory efficiency as the primary memory requirement for processing the data-stream. The memory requirement is directly proportional to the number of micro-clusters [13, 14]. OCED uses KDDCUP'99 data-stream with the same parameter setting as in Sect. 3.1 to compute the storage requirement.

The quantity of micro-clusters is recorded in OCED, CEDAS, and BOCEDS at an interval of 50 K data points and compared in Fig. 4. From Fig. 4, the micro-cluster count in BOCEDS is much higher than CEDAS as BOCEDS stores outliers in addition to micro-clusters for future reference, and a micro-cluster remains alive for more time than in CEDAS. The number of micro-clusters in

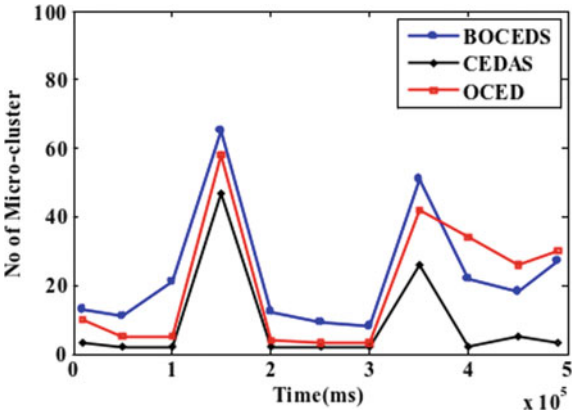


Fig. 4 Memory uses for clustering of KDDCUP'99 data-stream: the change in the number of micro-clusters in the current system

OCED is less than BOCEDS as OCED stores only micro-clusters. However, the amount of alive micro-clusters in OCED is higher than CEDAS as OCED uses the same procedure as in BOCEDS. Thus, OCED is more memory efficient than BOCEDS.

4 Conclusion

This paper includes OCED, a fully online approach for clustering of evolving data-stream in the domain of density-based clustering. Unlike other density-based clustering approaches, OCED learns the minimum micro-cluster radius and updates the radius towards optimal in a controlled way. These facts eliminate the requirement of predefining micro-cluster radius and grow micro-cluster in dense regions. Moreover, a procedure has been introduced for merging micro-clusters to reduce the unnecessary micro-clusters for improving memory efficiency as well as computational time. The experimental result demonstrates the ability of OCED in generating high-quality and arbitrary-shaped clusters occupying moderate storage. OCED shows its superiority over aligned clustering approaches concerning processing time and scalability. OCED is able to handle only numerical attributes. Future research could extend OCED to extract clusters in data-stream with mixed attribute types.

References

1. Su N, Liu J, Yan C, Liu T, An X (2017) An arbitrary shape clustering algorithm over variable density data streams. J Algorith Comput Technol 11(1):93–99
2. Nguyen H-L, Woon Y-K, Ng WK (2015) A survey on data stream clustering and classification. Knowl Inf Syst 45(3):535–569

3. Ramirez-Gallego S, Krawczyk B, Garcia S, Wozniak M, Herrera F (2017) A survey on data preprocessing for data stream mining: current status and future directions. Neurocomputing 239:39–57
4. Amini A, Wah TY, Saboohi H (2014) On density-based data streams clustering algorithms: a survey. J Comput Sci Technol 29(1):116–141
5. Zubaroğlu A, Atalay V (2021) Data stream clustering: a review. Artificial Intell Rev 54:1201–1236
6. Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. In: International conference on knowledge discovery and data mining, vol 96, pp 226–231
7. Cao F, Estert M, Qian W, Zhou A (2006) Density-based clustering over an evolving data stream with noise. In: Proceedings of the 2006 SIAM international conference on data mining, SIAM, pp 328–339
8. Mousavi M, Khotanlou H, Bakar AA, Vakilian M (2020) Varying density method for data stream clustering. Appl Soft Comput 97:106797
9. Ruiz C, Menasalvas E, Spiliopoulou M (2009) C-denstream: using domain knowledge on a data stream. In: International conference on discovery science, pp 287–301. Springer
10. Baruah RD, Angelov P (2012) Evolving local means method for clustering of streaming data. In: 2012 IEEE international conference on fuzzy systems, pp 1–8. IEEE
11. Baruah RD, Angelov P (2013) Dec: Dynamically evolving clustering and its application to structure identification of evolving fuzzy models. IEEE Trans Cybern 44(9):1619–1631
12. Hyde R, Angelov P (2015) A new online clustering approach for data in arbitrary shaped clusters. In: 2015 IEEE 2nd international conference on cybernetics (CYBCONF), pp 228–233. IEEE
13. Islam MK, Ahmed MM, Zamli KZ (2019) A buffer-based online clustering for evolving data stream. Inf Sci 489:113–135
14. Hyde R, Angelov P, MacKenzie AR (2017) Fully online clustering of evolving data streams into arbitrarily shaped clusters. Inf Sci 382:96–114
15. Rathore P, Kumar D, Bezdek JC, Rajasegarar S, Palaniswami M (2020) Visual structural assessment and anomaly detection for high-velocity data streams. IEEE Trans Cybern
16. Islam MK, Ahmed MM, Zamli KZ (2019) I-codas: an improved online data stream clustering in arbitrary shaped clusters. Eng Lett 27(4)
17. Steinhaus H (1999) Mathematical snapshots. Courier Corporation

Improving Sentiment Classification for Large-Scale Social Reviews Using Stack Generalization

Lamees Abu Romman, Shaheen Khatoon Syed, Majed Alshmari, and Md Maruf Hasan

Abstract The problem of identifying sentiment from customers' reviews has been an important issue for many years. Previously, different machine learning methods have been utilized to automatically categorize users' reviews into polarity levels such as positive, negative, or neutral. However, these methods suffer from low accuracy and recall. This paper presents an ensemble learning method using stacking generalization to build an accurate model for predicting sentiment polarity from social reviews. The basic concept of stacked generalization is fusing the output of a first-level classifier with a second-level classifier in a stacking manner. The diversity among the base classifiers with different features and weight measures is investigated in two domains (Twitter and Amazon product reviews), which provides a space for improving sentiment classification performance. Four types of singular classifiers: namely, support vector machine, boosted decision tree, Bayes point machine, and averaged perceptron, are used to build a two-staged and stacking model. The performance of singular and two-staged classifiers is compared with the proposed stacking model. The experiment results demonstrate that the stacking model outperforms the singular and two-staged classifiers on both datasets in terms of accuracy, precision, recall, and F1-score.

Keywords Sentiment analysis · Ensemble · Stacking · Combined-classifier · Base-learner · Meta-learner · Heterogeneous classifier

1 Introduction

The textual data on the Web and social media in the form of reviews has become a critical factor in purchasing decision-making. To use the embedded knowledge effectively, polarity detection to automatically categorize customers' reviews into a

L. Abu Romman (✉) · S. K. Syed · M. Alshmari · M. M. Hasan
College of Computer Science and Information Technology, Department of Information Systems, King Faisal University, Al Ahsa, Saudi Arabia
e-mail: laburomman@kfu.edu.sa

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_11

positive, negative, or neutral has become essential. Many studies have either used a lexicon-based approach or machine learning to identify sentiments from customers' reviews [1–3].

In lexicon-based approaches, a sentiment lexicon consists of some words and phrases, which can express positive or negative sentiments, which have been used to detect sentiments [4]. However, using only a sentiment lexicon for sentiment classification is not sufficient to accurately predict the semantic orientation of input documents due to the opposite orientations of the same words in different domains.

In machine learning-based approaches, different supervised learning techniques such as Naïve Bayes (NB), Support Vector Machine (SVM), and Decision Tree (DT) have been used to improve sentiment classification [5]. One major task of machine learning is to build an accurate predictive model for a specific domain using supervised learning [6]. A classifier is a model that classifies instances from a target dataset and assigns these instances to one of a predefined class. The model generally takes annotated polarities from a specific domain as sentiment categories and uses machine learning to build a sentiment classifier [6–8]. However, most studies either used singular or multi-stage classifiers, ignoring the efficient way of integrating different classifiers, which could improve the sentiment classification performance. On the one hand, which classification algorithm performs better among several algorithms remains a matter of debate. On the other hand, different classification algorithms are domain-dependent.

This study demonstrates the use of multi-stage classifiers in an efficient way in order to overcome the drawbacks of individual classifiers and benefit from each other's merits, to enhance the sentiment classification performance. We aimed to answer the question: "can an ensemble classifier model achieve performance better than a singular or two stages classifier?". We built an Ensemble learning classifier by combining a number of singular classifiers, such as SVM, Boosted Decision Tree (BDT), Bayes Point Machine, and Averaged Perceptron. To construct an ensemble system, many parallel individual classifiers are generated in a stacking manner, and then the output of base learners is fused using a meta-level classifier to aggregate the output [9]. The predictions from each base-level classifier are given as input for a meta-level classifier to predict the final class label [10]. Stacking generates an ensemble learner composed of heterogeneous classifiers, unlike other ensemble generation algorithms as bagging or boosting, which generate an ensemble of classifiers using homogeneous ensembles. In the stacking method, the resulted model is predicted to be more accurate than the singular model. Building an ensemble classifier aims to build a predictive model that combines multiple classifiers to obtain a classifier that outperforms stand-alone ones. However, the appropriate combination and number of classifiers in building stacking models remain a matter of debate.

Many studies have been conducted to select the appropriate combination of a base learning algorithm to build the ensemble learners. For instance, Džeroski and Ženko [11] proposed two extensions of stacking methods: utilizing an extended set of meta-level and the other by employing a multi-response model to learn from the meta-level. The results showed that using a multi-purpose extension model

outperforms the other stacking approaches. Similarly, Chalothom and Ellman [12] experimented with sentiment analysis of tweets to show how ensemble learning can improve performance. SVM, Naive Bayes, SentiStrength, and Stacking were used to form the ensemble learning, and majority voting is used to aggregate the output. The results showed that the combination of these classifiers achieved the best scores for both tweets and SMS datasets.

The goal of this work is to improve sentiment classification for large-scale social reviews using stack generalization. Four heterogeneous classifiers are selected to build an ensemble: SVM, BDT, Bayes point machine, and averaged perceptron. Each base classifier was evaluated individually before creating an ensemble classifier. The stacking technique is applied for constructing an ensemble of classifiers. Scored probabilities of training and testing dataset from each of the individual classifiers are used as features for building ensemble model using boosted decision tree. The results showed significant performance improvement in terms of accuracy, precision, and recall as compared to the singular and two-staged classifiers.

The paper is organized as follows. The literature review is presented in Sect. 2, highlighting the most popular and early studies conducted in the area of sentiment analysis. The proposed methodology is described in Sect. 3. In Sect. 4, experimental results are presented and compared with related work. Finally, the conclusion and implications are provided in Sect. 5.

2 Related Work

Many studies in the area of sentiment analysis used singular classifiers to predict sentiment class from social reviews. The most popular and early study in this area is conducted by Pang et al. [13], where different machine learning algorithms were applied to the movie reviews dataset. The results proved that machine learning techniques outperformed human-produced baselines recall scores ranged from 50 to 69%. Similarly, Prasad et al. [14] classified user opinions in the Indian language using decision tree algorithms. Experiments were performed on the training dataset with a balanced number of positive, negative, and neutral tweets. The results showed a better performance of C4.5 (DT) with an average f-score of 0.804. Bilal et al. [15] performed polarity mining on public opinions written in Roman-Urdu language using supervised and unsupervised approaches. The opinions were extracted from a blog. Three classification models (Naive Bayes, J48 (DT), and KNN) were used for text classification. The results showed that Naive Bayes outperformed KNN and DT and achieved better accuracy, recall, precision, and F-measure. Altawaier and Tiun [16] identified a functional approach to apply sentiment analysis on Arabic tweets. Three classification techniques such as SVM, NB, and DT, were used due to their effective capability to deal with text categorization, where the number of features is huge. The results showed that DT outperformed NB and SVM with 78% F-scores.

Besides singular classifiers, many studies used combined multiple supervised classifiers to achieve higher accuracy. For instance, Catal and Nangir [1] proposed a multi-classifier system for sentiment classification on Turkish reviews in three different domains, i.e., book, shopping, and movie. The majority voting algorithm was used on top of Naive Bayes and SVM. The proposed approach showed better performance than Naive Bayes and SVM individually and showed improvement in the effectiveness of sentiment analysis with an accuracy 86.13%. Similarly, Chalothom and Ellman [12] used the ensemble learning method with majority voting on top of SVM, Naïve Bayes, and SentiStrength. Tweets and SMS datasets were used to evaluate the proposed system. The overall best score for both datasets was achieved by combining the classifiers using majority voting. Sudheep Elayidom and Idikulla [10] performed an ensemble technique to answer, "do classifier performance will be improved by employing the stacking method?". They focused on how heterogeneous classifiers can be stacked to improve the performance of the classifier. Three classifiers, decision tree, neural network, and Naive Bayes, were selected to build an ensemble. These classifiers were evaluated individually to compare their results with the ensemble. They observed that the stacking ensemble method achieved better accuracy than singular classifiers. Combining classifiers with stacking achieved an accuracy of 82%, which was better than other approaches. Another relevant study was proposed by Džeroski and Ženko [11], where an experiment was conducted to answer the same question, "whether combining classifiers with stacking better than selecting the best one?". Two extensions were proposed, where the first one is utilizing a set of meta-level features and the second one using a multi-response model to learn from the meta-level. The results showed that multi-response outperformed the other approaches. Another study Alnashwan et al. [17] proposed an ensemble learning approach by employing four base learners, i.e., SVM, Bayes point machine, logistic regression, and decision forest. Three different labeled datasets were used to evaluate the effectiveness of the proposed approach. The results showed that combining the learners minimizes the error rate as compared to the individual classifiers. Saleena, [18] proposed an ensemble classifier by combining Naïve Bays, logistic regression, random forest, and SVMs into a single classifier, with an aim to improve the performance of the sentiment classification. The proposed classifier performed better than single classifiers and majority voting ensemble classifier.

In this study, we proposed two types of multiple supervised learners using ensemble-learning methods. The first proposed classifier is a super classifier using stack generalization, where four individual classifiers are combined to build a super classifier using the ensemble method. The second is two-staged classifiers, which built six classifiers by combining every two classifiers together. The performance results for the super classifier, the two-staged classifiers, and singular classifier is compared using accuracy, precision, recall, AUC, and F1-score on two types of datasets, i.e., tweets from Twitter and reviews from Amazon.com.

3 Methodology

In this work, we have developed an end-to-end solution to facilitate and gain intelligence for business-related social review. The proposed methodology consists of two main modules, including (1) the development of a tool to automate the process of creating labeled example sets. (2) development of supervised algorithms to predict sentiment polarity of new reviews. The overall architecture of the proposed methodology is shown in Fig. 1. In the following section, we describe the detail of each module.

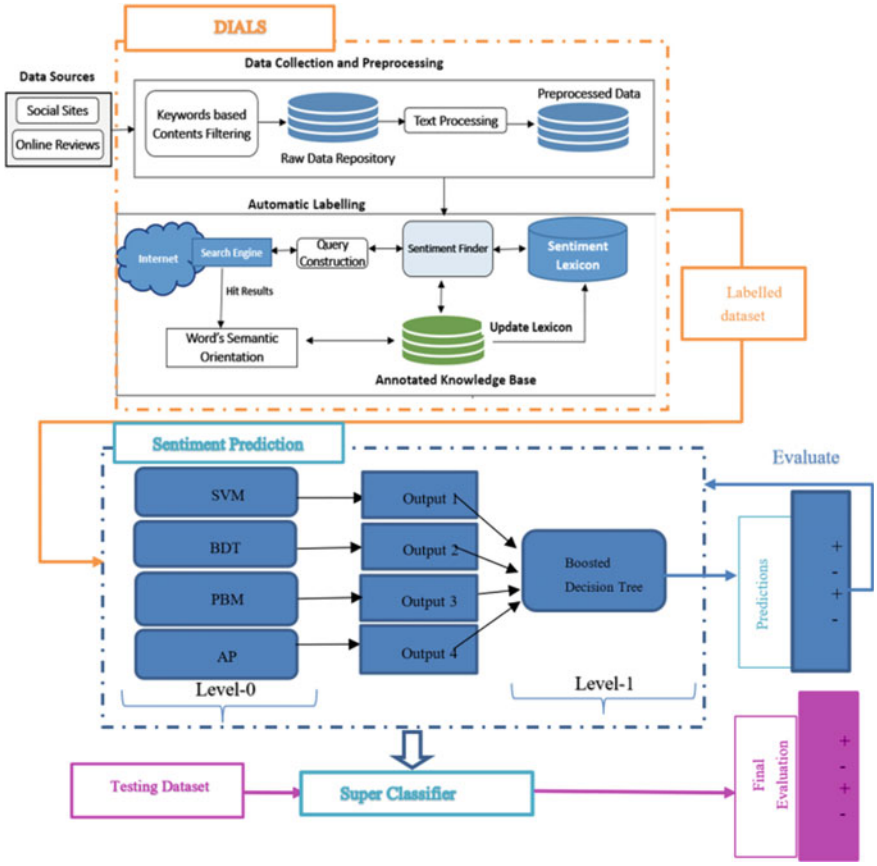


Fig. 1 The architecture of the proposed system for predicting sentiment polarity from social reviews

3.1 Development of Labelled Dataset

To automate the manual process of creating labeled example sets, we developed a tool called Domain-Independent Automatic Labeling System (DIALS). As shown in Fig. 1, DIALS takes reviews from any social site and creates a labeled dataset that is used to train supervised classification algorithms in the next step. First, the system uses an initial dictionary with known semantic orientation (SO) to calculate the semantic orientation for each word in the given text. If any word in the given text is not found in the lexical database, the system utilizes an online search to calculate the co-occurrence of target words with positive and negative contextual words to calculate the semantic orientation. An iterative process continues until the semantic orientation of all words' from the target text is calculated. For the detailed implementation of DIALS, see our previous work [19]. The labeled dataset is then used to train the classifiers discussed in the next step.

3.2 Proposed Sentiment Classification Model

This section introduces our proposed sentiment classification model and presents individual classifiers used in building the model. In order to perform sentiment classification, we built a super-classifier based on the stacking ensemble learning method, according to Dietterich [20]. The main reasons to select the ensemble learning is (1) it reduces the chance of selecting the wrong class; (2) as some of the learning algorithms are relied on a local search, utilizing ensemble learning, running of the local search starts with many classifiers, which will avoid the local search stuck and achieve better performance than individual classifiers; (3) an ensemble learning expands the representation space to provide a proper target function. We built several classifiers using two different datasets to identify the best binary classification based on the performance results. In the stacking ensemble technique, scored probabilities are used from each of the individual classifiers, and boosted decision tree is used over the scored probabilities to predict the class label, as shown in Fig. 1. The scored probabilities are taken from the training and testing dataset and used in a Two-Class Boosted Decision Tree by tuning its hyper-parameters using the same method that is used for individual classifiers. Hence, the Super classifier takes as input from four features and adding scored probabilities coming from the individual classifier. The novelty of the super classifier comes from its ability to use the ensemble method in addition to using parameters optimization on the individual classifiers and on the super classifier. The following algorithm illustrates the details of the Super Classifier.

Algorithm (1) Super Classifier algorithm
Input: LaT_i = Labeled texts
LEARNER = [Support Vector Machine, Boosted Decision Tree, Bayes Point Machine, Averaged Perceptron, Super Classifier]

Output: Classified text

```
For L in LEARNER
  Repeat N Times
    For data in LaT_i
      TRAIN = random 80% of data
      TEST = data – TRAIN
      Predictor Train LEARNER with TRAIN
Accuracy = PREDICTOR on TEST
Precision = PREDICTOR on TEST
Recall = PREDICTOR on TEST
      AUC = PREDICTOR on TEST
      F1- score = PREDICTOR on TEST
    END
  END
```

Following individual classifiers are used to build ensemble classifier:

- Two-Class Support Vector Machine (SVM) tries to find the maximum margin hyperplane, which provides the maximum separation between classes, and the samples that are closest to this hyperplane are known as support vectors [21].
- Two-Class Boosted Decision Tree (BDT) uses an ensemble of decision trees by applying the boosting concepts. In the DBT approach, the second tree corrects the error of the first tree, and the third tree corrects the error of the second tree, and so on. The predictions are based upon the ensemble of trees produced in this fashion until one of the criteria is reached [22].
- Two-Class Bayes Point Machine (BPM) approximates the theoretically optimal bayesian average of linear classifiers (in terms of generalization performance) by choosing one "average" classifier, the Bayes Point. Because the Bayes Point Machine is a Bayesian classification model, it is not prone to overfitting to the training data. It does not require parameter tuning and does not need data to be normalized [17].
- Two-Class Averaged Perceptron (AP) is a simple version of a neural network, which classifies possible outputs based on a linear function. The outputs are then combined with a set of weights that are derived from the feature vector [23].

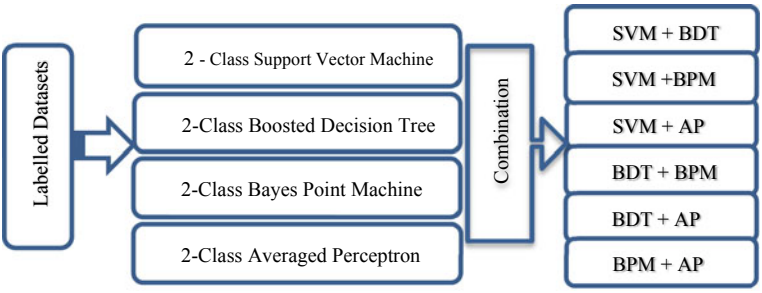


Fig. 2 Binary combination of singular classifiers

We also built a set of binary combinations of singular classifiers, where each classifier is built by combining two algorithms from the above-selected algorithms. We built six classifiers with the following arrangements: SVM + BPM, SVM + BPM, SVM + AP, BDT + BPM, BDT + AP, BPM + AP. The arrangement is shown in Fig. 2. The result of singular and combined classifiers are compared with the proposed ensemble learning model and discussed in Sect. 4.

4 Experimental Results

We used Twitter and amazon reviews to validate the proposed methodology. The dataset detail is shown in Table 1. To prepare tweets dataset, crawled tweets related to Saudi Telecom companies using several keywords such as STC_KSA, STCcare, ZainKSA_ Mobily, Mobily 1100, MobilyBusiness. The second dataset was extracted from Amazon product reviews using a handy tool, which is a part of the Helium 10 chrome extension. After basic pre-processing, both datasets were labeled using DIALS tool. For more detail on labeling using DIALS, see our previous work Khatoon et al. [19]. The labeled datasets are then used to conduct experiments for validating singular, combined, and super-classifier.

A two-dimensional confusion matrix is used to assess the accuracy of the sentiment classifiers by showing the number of actual and predicted labels. The confusion matrix consists of four values: true positives (TP) manually and automatically classified as positive), false negatives (FN) (automatically classified as negative), false positives (FP) only automatically classified as positive and true negative (TN) both manually and automatically classified as negative (see Table 2).

Table 1 Datasets selected for experimentation

Datasets	Source	Size	Language
tweets	Twitter	1201	English
reviews	Amazon.com	1028	English

Table 2 Confusion matrix

	Actual positive	Actual negative
Predicted positive	True positive (TP)	False positive (FP)
Predicted negative	False negative (FN)	True negative (TN)

The performance of the different model is evaluated by calculating accuracy, precision, recall and F1 on the training and testing datasets calculated as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + TN = FN} \tag{1}$$

$$Recall(R) = \frac{TP}{TP + FN} \tag{2}$$

$$Precision(P) = \frac{TP}{TP + FP} \tag{3}$$

$$F1 = \frac{2PR}{P + R} \tag{4}$$

For Tweets, the number of labeled tweets were 1201; these tweets were pre-processed and split into 60% for training (721 tweets) and 40% for testing (480 tweet). For Amazon reviews, the number of labeled reviews were 1028, these reviews were pre-processed and split into 60% for training (618 reviews) and 40% for testing (411 reviews). The analysis was executed in the following three sets of experiments. The overall results are shown in Tables 3 and 4. All the experiments were performed using Python programming and relevant machine learning libraries.

In the first experiment, individual classifiers, i.e., SVM, BDT, Bayes point machine, and averaged perceptron, were tested on both datasets. For the tweets dataset, SVM outperformed the other learners by achieving 69% accuracy, 75% precision, 57% F1-score, and 47% recall. For the Amazon reviews dataset, BDT outperformed other classifiers by achieving 88% accuracy and precision, 90% for recall, and F-score.

In the second experiment, a set of binary combinations of base classifiers were built to form six combined classifiers. Each combined classifier was evaluated on both datasets to compare the performance. The best combined classifier using tweets was Boosted decision tree + Bayes point machine with 70% accuracy, 62% precision, 75% recall, and 68% F1 score. This binary combination of Boosted decision tree + Bayes point machine also achieved higher performance on amazon reviews dataset with 90% accuracy, 91% precision, 98.5% recall, and 95% F1 score.

In the third experiment, the super classifier is built by employing the stacking ensemble method. Super classifier showed better performance compared to the singular and combined classifiers. For the tweets dataset, it classified 337 tweets correctly and 143 incorrectly and achieved 78% for both accuracy and precision.

Table 3 Performance evaluation of singular, multi-base level binary classifiers, and the super classifiers on Tweet dataset

Classifiers (Using Tweets)	#Correct classifications	#Incorrect classifications	Evaluation measures				
			Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Singular base classifier							
BDT	319	151	67	73	42	53	72
AP	318	162	67	71	43	53	73
SVM	323	157	69	75	4	57	75
BPM	314	166	65	66	45	55	64
Multi-base level binary classifiers							
Multi-base level binary classifiers	101	91	53	42	31	34	44
BPM + SVM	104	88	54	44	31	36	49
BDT + SVM	104	88	54	44	32	37	50
BDT + BPM	135	57	70	62	75	68	80
AP + BDT	101	91	53	43	39	41	47
AP + BPM	103	89	54	44	37	40	47
Super classifier							
Ensemble (Super classifier)	337	143	78	78	79	73	82

Table 4 Performance evaluation of singular, multi-base level binary classifiers, and the super classifiers on Amazon reviews

Classifiers (Using Amazon reviews)	#Correct classifications	#Incorrect classifications	Evaluation measures				
			Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC (%)
Singular base classifier							
BDT	352	59	88	88	90	90	83
AP	330	81	80	86	89	85	80
SVM	351	60	95	86	90	89	83
BPM	344	67	79	82	81	80	64
Multi-base level binary classifiers							
AP + SM	331	80	80	86	93	89	37
BPM + SVM	301	110	73	84	85	84	22
BDT + SVM	288	123	70	83	81	82	22
BDT + BPM	370	41	90	91	98	95	89
AP + BDT	316	95	77	83	91	87	22
AP + BPM	326	85	79	84	94	88	24
Super classifier							
Ensemble (Super classifier)	375	36	94	92	99	96	89

Also, super classifiers achieved 79%, 70%, and 73% for recall, F1-score, and AUC, respectively. The Super classifier achieved higher performance as compared to the singular and combined classifiers. It classified 37 reviews correctly and 36 incorrectly and achieved 94% accuracy and 92% precision. Also, the super classifier outperformed the other two methods with 99%, 96%, and 89% of recall, F1-score, and AUC, respectively. According to the performance results, employing a super classifier improved the sentiment classification.

One challenge of this study is that there is no global best solution exists to verify why one algorithm performs better than the other. Selecting the number of base classifiers is another challenge. Previous studies Sudheep Elayidom and Idikulla [10] used 3 to 7 base classifiers based on different application types.

We compared the results with a similar study conducted by Catal and Nagir [1], where they used the ensemble method to improve the results of multiple classifiers. They used a majority voting algorithm in conjunction with Naïve Bayes, SVM, and Bagging parameters of SVM. The performance of the voting algorithm is evaluated using three different datasets (booking, shopping, and movies datasets). Labeled reviews were used to train the proposed classifier utilizing a lexical-based approach to categorize reviews into positive or negative. Experimental results showed that multiple classifier systems achieved an accuracy of 86%, which is very high as compared to the accuracy of individual classifier. However, the super classifier's accuracy, which is proposed in this study, is 94%, which is 8% higher than the multi-classifier proposed by Catal and Nagir [1]. The obvious improvement in our work is the automatic labeling system, which is not only using the available lexicon but finding the polarity of the unknown words using web-search, unlike Catal and

Nagir work, where they only used available lexicon. The other reason for high performance in our case is the different combination of classifiers, which proven to be very efficient for Twitter data.

The second study we compared is proposed by Wan and Gao [24], where they built an ensemble approach by integrating five popular classification methods: namely, Naïve Bayes, SVM, C4.5 decision tree, Bayesian network, and random forest classifier. The results showed that the proposed ensemble approach outperformed the individual classifiers in the airline service Twitter dataset, where the dataset was labeled manually into two categories. The ensemble approach increased the sentiment classification performance by 91.7% and the lowest error rate among other classification approaches. The results achieved by their ensemble approach are very close to the result of this study. However, in this study, manually labeling is done automatically, and datasets from different domains are used.

Chalothom and Ellman [12] used ensemble learning using majority voting by combining SVM, naïve Bayes, SentiStrength, and stacking. Two datasets, tweets, and SMS were used to evaluate the performance. F score was used to measure the efficiency of the ensemble learner with 86.05% for tweets and 88.8% for SMS. Our work and other related works showed that the stacking method is better suited for combining base types of classifiers to achieve the best classification performance.

5 Conclusion and Future Work

In this research, we have developed an end-to-end solution for gaining business intelligence from social reviews. We developed a purely supervised and domain-independent labeling system called DIALS to automate the process of labeling large-scale social data. In this work, on top of DIALS, we built different classification approaches for sentiment analysis. Overall, the system takes a labeled dataset from DIALS to create a large training corpus for subsequently training classification models for classifying social data into positive and negative classes. We built a combined classifier called super classifier by fusing the output of the first-level meta classifier with the second-level meta classifier in a stacking manner. Four types of base-level classifiers, i.e., SVM, boosted decision tree, Bayes point machine, and averaged perceptron, are used to build the stacking model. Experiments are conducted to validate the performance of the super classifier using datasets from Twitter and Amazon product reviews. The results of the super classifier are compared with singular and multi-base level binary classifiers using accuracy, precision, recall, F1-score, and AUC.

Results showed that super-classifier has proven consistently effective for both datasets and achieved higher performance compared to the singular and combined classifiers. For the tweets dataset, the super-classifier achieved 78% for both accuracy and precision. For the Amazon product review dataset, the super-classifier outperformed the other two approaches with 99%, 96%, and 89% in terms of recall, F1-score, and AUC, respectively.

Although results on two datasets conclude that stack generalization using four base-level classifiers performs better than singular classifiers, however, choosing the right number of base classifiers, as well as learning levels, is still a challenge. In future research, we will increase the base classifiers and learning levels and validate the super-classifier on the dataset from multiple domains. There are many challenges in the ensemble classification field that are worth further exploring in the future. Specifically, ensemble classification models contain many characteristics like diversity, accuracy, generalization and so on, and these characteristics are conflicting in improving performances of the models in certain cases. Hence, there are many existing methods exploring to combine these characteristics using mixture models or multi-objective functions to optimize them simultaneously. Ensemble learning techniques may be used to alleviate significant challenges to machine learning algorithms. Some of these challenges are class imbalance problems, where one class has more examples than other classes. In such cases, ensemble methods may be applied in a way that reduces the class imbalance problem.

Moreover, the distribution of features and the labels tend to change over time, which repeatedly influences the performance of the model over time. Ensemble-based approaches can serve as a repair for this problem. Another challenge is the curse of dimensionality to a large number of features, which causes exponentially increases the search space, and model's overfitting. This challenge can be resolved by specific ensemble learning methods to minimize the impact of this phenomenon.

6 Funding Statement

The authors extend their appreciation to the Deputyship for Research and Innovation, Ministry of Education in Saudi Arabia, for funding this research work through project number 523.

References

1. Catal C, Nangir M (2017) A sentiment classification model based on multiple classifiers. Appl Soft Comput 50:135–141
2. Ji Y, Wu W, Chen S, Chen Q, Hu W, He L (2017) Two-stage sentiment classification based on user-product interactive information. Knowl Based Syst 203:106091
3. Sharma A, Dey S (2013) A boosted SVM based sentiment analysis approach for online opinionated text. In: Proceedings of research in adaptive and convergent systems, pp 28–34
4. Liu B (2012) Sentiment analysis and opinion mining. Syn Lect Hum Lang Technol 5(1): 1–167
5. Xia R, Zong C, Li S (2011) Ensemble of feature sets and classification algorithms for sentiment classification. Inf Sci 181(6):1138–1152
6. Aggarwal CC (2015) Data mining: the textbook. Springer
7. Zhou Z-H (2012) Ensemble methods: foundations and algorithms. CRC press

8. Sesmero MP, Ledezma AI, Sanchis A (2015) Generating ensembles of heterogeneous classifiers using stacked generalization. Wiley Interdiscip Rev: Data Min Knowl Discov 5(1):21–34
9. Zhou Z-H (2009) Book ensemble learning, encyclopedia of biometrics. Springer Berlin, pp 270–273
10. Sudheep Elayidom M, Idikulla SM (2012) Design and development of data mining models for the prediction of manpower placement in the technical domain. Cochin University of Science and Technology
11. Džeroski S, Ženko B (2004) Is combining classifiers with stacking better than selecting the best one? Mach Learn 54(3):255–273
12. Chalothom T, Ellman J (2015) Simple approaches of sentiment analysis via ensemble learning: information science and applications. Springer, pp. 631–639
13. Pang B, Lee L, Vaithyanathan S (2002) Thumbs up?: sentiment classification using machine learning techniques. Association for Computational Linguistics, pp 79–86
14. Prasad SS, Kumar J, Prabhakar DK, Pal S (2015) Sentiment classification: an approach for Indian language tweets using decision tree. Springer, pp 656–663
15. Bilal M, Israr H, Shahid M, Khan A (2016) Sentiment classification of Roman-Urdu opinions using Naïve Bayesian Decision Tree and KNN classification techniques. J King Saud Univ Comput Inf Sci 28(3):330–344
16. Altawaier MM, Tiun S (2016) Comparison of machine learning approaches on Arabic Twitter sentiment analysis. Int J Adv Sci Eng Inf Technol 6(6):1067–1073
17. Alnashwan R, O'Riordan AP, Sorensen H, Hoare C (2016) Improving sentiment analysis through ensemble learning of meta-level features. Sun SITE Central Europe (CEUR)/RWTH Aachen University
18. Saleena N (2018) An ensemble classification system for twitter sentiment analysis. Procedia Comput Sci 132:937–946
19. Khatoon S, Romman LA, Hasan MM (2020) Domain independent automatic labeling system for large-scale social data using Lexicon and web-based augmentation. Inf Technol Control 49(1):36–54
20. Dietterich TG (2000) Book ensemble methods in machine learning. Springer, pp 1–15
21. Pavlidis P, Wapinski I, Noble WS (2004) Support vector machine classification on the web. Bioinformatics 20(4):586–587
22. Pal M (2007) Ensemble learning with decision tree for remote sensing classification. World Acad Sci Eng Technol 36:258–260
23. Krithika R, Narayanan J (2015) Learning to grade short answers using machine learning techniques. pp 262–271
24. Wan Y, Gao Q (2015) An ensemble sentiment classification system of twitter data for airline services analysis. IEEE, pp 1318–1325

A Comparative Study on Autism Among Children Using Machine Learning Classification

Ainie Hayati Binti Noruzman, Ngahzaifa Ab Ghani, and Nor Saradatul Akmar Zulkifli

Abstract Autism Spectrum Disorder (ASD) is a neurodevelopment that affects communication and behavior in humans. It is a condition associated with a complex brain disorder, leading to significant changes in a human being's social interaction and behavior. Typically to detect toddlers who have ASD through screening tests is very expensive and time-consuming. Typically, detecting toddlers who have ASD through screening tests is very expensive and time-consuming. However, with machine learning technology today, autism can be diagnosed efficiency and accuracy. This study aims to analyze and make a comparison on which prediction model that gives a high accuracy after the feature selection. The importance of attributes is investigated using correlation and the predictive models are constructed for the detection of this disorder in children. The dataset consists of 1054 instances and each instance includes 19 attributes. Experimental results clearly show that using feature selection with 10 attributes can lead the impact of accuracy with predictive model of Random Forest (RF) returns the highest accuracy with 94.78%. The findings also indicated that the number of questions in screening tools can be reduced and give an impact with the good results.

Keywords Machine learning · Classification · Feature selection

A. H. B. Noruzman (✉) · N. A. Ghani · N. S. A. Zulkifli
Faculty of Computing, University Malaysia Pahang, Pekan, Malaysia
e-mail: pcp20001@student.ump.edu.my

N. A. Ghani
e-mail: zaifa@ump.edu.my

N. S. A. Zulkifli
e-mail: saradatulakmar@ump.edu.my

A. H. B. Noruzman · N. A. Ghani
Centre for Software Development and Integrated Computing, University Malaysia Pahang, Pekan, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_12

1 Introduction

1.1 Autism Spectrum Disorder

ASD is the second most prevalent neurodevelopmental disorder among children, and its estimates vary markedly over time and among countries, depending on multiple epidemiological variables [1]. Due to the rising prevalence of autism, much research has centred on creating a new strategy for cost-effective and quickest tracking in diagnosing ASD. Traditionally, to determine the children have autism or not, clinical services and thorough evaluation are required for an accurate diagnosis by clinical experts. Children may undergo clinical methods such as Autism Diagnostic Interview-Revised (ADI-R) and Autism Diagnostic Observation Schedule Revised (ADOS-R). Both methods based on behavioural observation are often a lengthy process and high cost for an ASD diagnosis [2]. Therefore, providing a quick and effective solution can help significantly practitioners predict ASD traits at a much earlier stage.

Nowadays, applying science and advanced technology in curing patients is appealing due to rapid information technology (IT) changes. One supportive approach that can automate ASD screening and develop accuracy and efficiency is machine learning usage. Machine learning can speed up the prediction process and optimize the autism classification variables and decide whether a child has an ASD or not [3]. According to several studies, machine learning has contributed significantly to improving the diagnosis of ASD [2, 4–6]. Although there are already similar studies in classified ASD using machine learning, it still lacks adequate studies. Therefore, our contributions to this paper are focus on pre-processing using feature selection and classify.

This paper is organized as Sect. 2 describes the related works in machine learning, Sect. 3 explains the method and material, Sect. 4 methodology, Sect. 5 explains result and discussion and finally, Sect. 6 describes the conclusion of paper.

2 Related Works in Machine Learning

2.1 Feature Selection in Machine Learning

More than one is study considered as works necessary for the ASD classification using machine learning. Currently, machine learning is a fast growing trend in the health care industry [7]. The technology helps medical experts to analyze and predict the data for decision making and diagnostic purposes. One of the methods is Feature Selection. Feature Selection has often been applied to autism screening questionnaires to identify minimal feature subsets to speed up the diagnosis process [8]. For instances, Wall et al. [9] used machine learning to shorten the observation-based screening and diagnosis of autism. Their work proved that 8 of

the 29 items on Module 1 of the Autism Diagnostic Observation Schedule (ADOS) categorize autism with complete 100% accuracy. Kosmicki et al. [10] also indicated that 9 of the 28 items from Module 2 and 12 of the 28 items seized from Module 3 of ADOS are sufficient to detect ASD through feature-selection-based machine learning with 98.27% and 97.66% accuracy, respectively.

Work by Thabtah [11] and Abdullah et al. [12], has agreed that ML using LR classifier perform good accuracy. Work by Thabtah [11] used information gain (IG) and tested with autism screening of adults and adolescents only. The results produce approximately 87% sensitivity, accuracy, and specificity for adults but a lower accuracy of 85% for adolescents, which can be acceptable. The same model should be tested for children dataset in future works and compare for better improvement. Where else, Abdullah et al. [12] used the Least Absolute Shrinkage and Selection Operator (LASSO) as feature selection. The performance is improved as the number of features reduced from 19 to 13 features by using Chi-Square selected features, especially with Logistic Regression classifier present 100% accuracy. Another similar study by Erkan and Thanh [13] has experimented using ASD child dataset based on AQ10 and founds that RF gives 100% accuracy of ASD classification compared to k-NN and SVM. The data were divided randomly into training and test varying in 50, 60, 70, 80, 90%, and the remaining 100%. After splitting data into two parts: training and test, the data randomly selected 100 times to test the classification methods. Limitation research is small datasets been tested, which about 249 out of 262 after exiting the missing data.

Other work by Devika Varshini and Chinnaiyan [14] has experimented with data using simple pre-processing steps combined with suitable data encoding. Various classifier algorithms such as LR, KNN, and RF give rise to comparable results with state-of-the-art. The accuracy, F1 score, and efficiency are calculated based on the AQ-10 Adults and Toddlers datasets. The accuracy score results compared that KNN is 69.2% LR, and RF classifiers are 68.601% and 67.78%, respectively. Mahadev et al. [15] proposed a predictive model tested with the AQ10 dataset. The result found that using AdaBoost and RF can predict autism traits with 96.20%, 100.00% accuracy for age groups below three years, while other models cannot.

3 Materials and Methods

3.1 Data Set

The dataset for this research was compiled and publicly accessible at https://www.kaggle.com/fabdelja/autism-screening-for-toddlers [16]. In addition, the website also explicitly notes that the use of data is for research purposes only and advises users of the use of data. The detailed description of the data set is summarized in Table 1.

Table 1 ASD toddlers summary

Attribute Id	Attributes	Attributes and data types description
1	Name dataset	Toddler dataset July 2018.csv
2	Data type	Predictive and descriptive: nominal/categorical, binary and continuous
3	Attribute type	Categorical, continuous and binary
4	Missing values	No missing value
5	Number of instances	1054
6	ASD class	Yes/No

Table 2 List of attributes in dataset

Attribute Id	Attributes	Attributes and data types description
0	Case no	Case number (Number)
1–10	A1–A10 score	Screening question (Binary (0, 1))
11	Age_Mons	Patient age by months Number)
12	Qchat-10-Score	Screening score (Integer)
13	Sex	Gender: Male or female (String)
14	Ethnicity	Nationality (String)
15	Jaundice	Patient have a Jaundice by birth (Number)
16	Family_mem_with_ASD	Any members of family who experienced with ASD before (String)
17	Who completed the test	Person completed the test (String)
18	ASD/Class Traits	ASD/Non ASD class (String)

These datasets have 19 common attributes that are used for prediction. These attributes are listed shown in Table 2.

Figure 1 shows the visualization of attributes indicate class/ASD traits labels of binary classification which is Yes and No. The different colors show the total number of instances indicate that the Blue color represent of 'No' and the Red color represent 'Yes'.

4 Methodology

The framework of the proposed methodology is illustrated in Fig. 2. It consists of five phases: Data collection, Data Cleaning, Developing a prediction model, Test Model and Evaluating the prediction model accuracy.

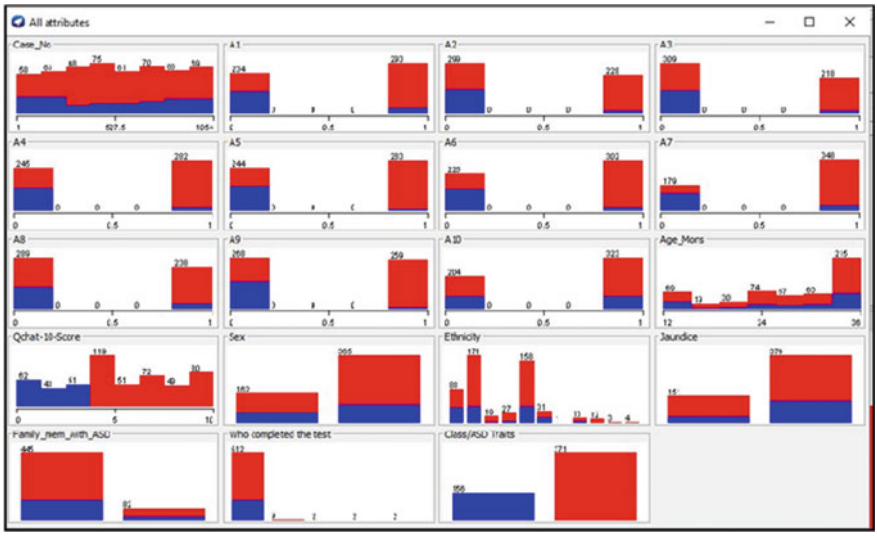


Fig. 1 Visualizations of attributes

4.1 Proposed Methodology

See Fig. 2.

4.2 Data Collection

Toddlers Dataset that has 1054 instances to be as an input.

4.3 Data Cleaning and Pre-processing Using WEKA

In this step, we used WEKA (Waikato Environment for Information Analysis Version 3.8.3) to clean and pre-process data. Since there are no missing values in the dataset, the Feature Selection methods have been used to identify the best feature to test with the prediction model.

The Feature Selection used the Attribute Evaluator *(CorrelationAttributeEval)* and Search (*Ranker*) to measure the correlation between attributes and to determine the positive correlation value of the attribute in the class. Any score from +0.4 to +1 indicates a moderate correlation since the attributes are increase at the same time [17]. Figure 3 show the effect of Feature Selection using WEKA.

The attribute ranked from 0.7958 to 0.3462 indicates that the attributes number 13, 6, 8, 10, 7, 5, 2, 3, 4 and 9 are identified as moderate correlation and selected. The ranked attributes below and equal 0.1 are not to selected in this study.

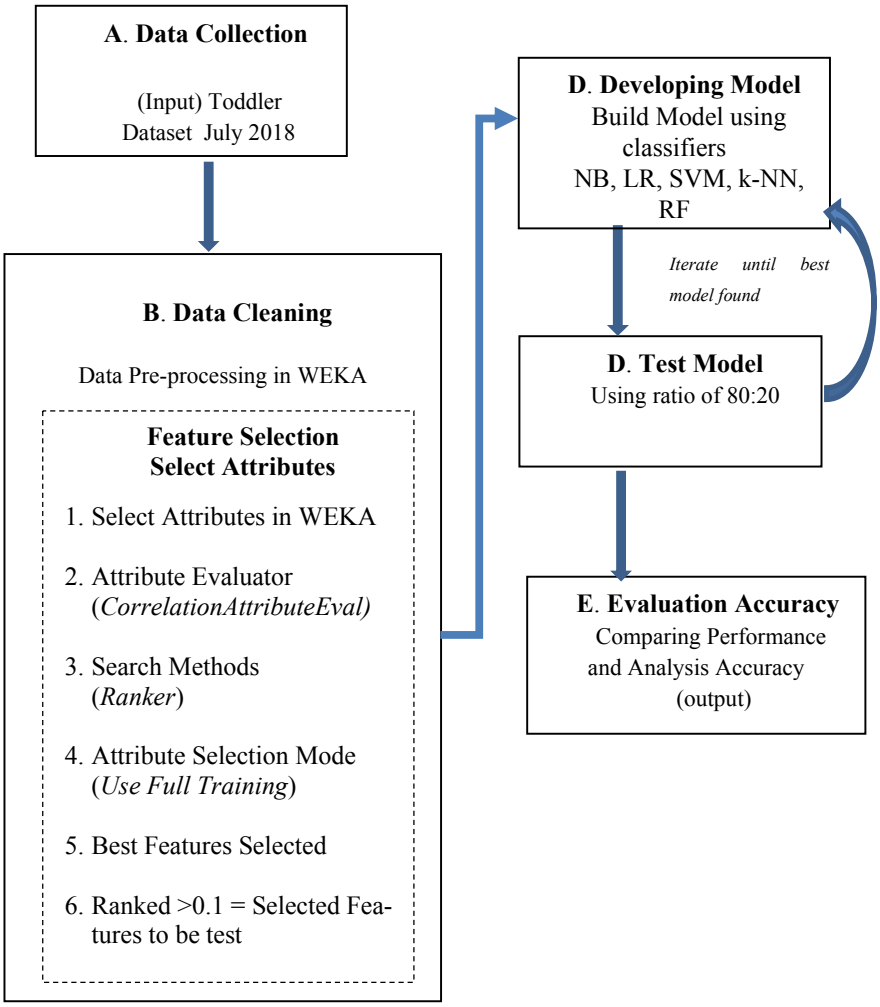


Fig. 2 The proposed methodology

However, the 'Qchat-10-score' attribute is removed for this analysis in order to test the selected attributes may or may not provide meaningful prediction to the dataset.

4.4 Developing Model

In this step we used WEKA in order to classify the datasets. The most popular machine learning methods were applied for the classification problem is the

```
=== Attribute Selection on all input data ===

Search Method:
        Attribute ranking.

Attribute Evaluator (supervised, Class (nominal): 19 Class/ASD Traits ):
        Correlation Ranking Filter
Ranked attributes:
 0.7958  13 Qchat-10-Score
 0.5733   6 A5
 0.5618   8 A7
 0.546   10 A9
 0.5412   7 A6
 0.504    5 A4
 0.4746   2 A1
 0.4236   3 A2
 0.4096   4 A3
 0.3462   9 A8
 0.1503  11 A10
 0.0892  16 Jaundice
 0.0854  18 Who completed the test
 0.0815  14 Sex
 0.0719  15 Ethnicity
 0.0437   1 Case_No
 0.0384  12 Age_Mons
 0.0146  17 Family_mem_with_ASD

Selected attributes: 13,6,8,10,7,5,2,3,4,9,11,16,18,14,15,1,12,17 : 18
```

Fig. 3 Ranked attributes in WEKA

k-Nearest Neighbour method (kNN), the Support Vector Machine method (SVM), the Naïve-Baiyes method (NB), the Logistic Regression method (LR) and the Random Forests method (RF).

4.5 Testing Model

80% of the data is used to train the model, and 20% is used to test. Figure 4 shows the final training, testing and validation sets on which classification has been performed.

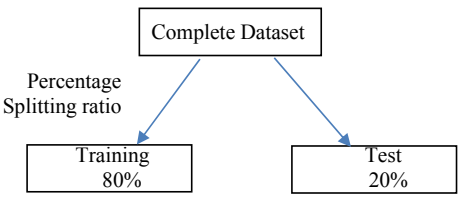


Fig. 4 Training and testing data

Table 3 ASD toddlers summary

Complete dataset	Training	Test
1054	843	211

Table 3 shows the number splitting instances when the whole dataset has been split into two parts with a ratio of 80:20 respectively.

4.6 Evaluation Accuracy

The classification performance is calculated by counting the accuracy, build time, correctly classified, incorrectly classified, recall and precision.

5 Result and Discussion

Experimental results of various machine learning algorithms approach with all features selection have been shown for ASD screening data for Toddlers. 10 features are selected to find the specificity, sensitivity, and accuracy of the predicted model. Table 4 shows the model with parameter using WEKA.

The overall performance measures of all machine learning classifiers with dataset have been shown in Table 5.

Based on the results depicted in Table 5, it is inferred that classification of ASD traits, the RF algorithms perform well as compared to the NB, SVM, LR and K-NN. It perceived that the RF algorithm gives the most accurate model to predict ASD traits. The RF has 200 of correctly classified with the lowest incorrectly classified of 12 showing that RF algorithms of is higher than others which is 94.78% accuracy compared to the k-NN with slightly lower 94.31%. Meanwhile, accuracy produced by LR and SVM shows the equally results with 93.83% and 93.36%, respectively. The lowest results is algorithms NB with 91.94%. This means that with training and test with ratio of 80:20 using a RF algorithms able to produce the best accuracy result compared than other predictive models. This accuracy rate performed indicates that the attributes in the dataset can be reduced and give the good results.

Table 4 Model parameter using WEKA

Predictive model	Parameter
NB	weka.classifiers.bayes.NaiveBayes
SVM	weka.classifiers.functions.SMO
LR	weka.classifiers.functions.Logistic
RF	weka.classifiers.trees.RandomForest
k-NN	weka.classifiers.lazy.IBk

Table 5 Accuracy of predictive model

Predictive model	Accuracy %	Build time (s)	Correctly classified	Incorrectly classified	Recall	Precision
NB	91.94	0	194	17	0.910	0.979
SVM	93.36	0.03	197	14	0.949	0.961
LR	93.83	0.03	198	13	0.962	0.955
RF	94.78	0.02	200	11	0.955	0.974
k-NN	94.31	0.14	199	12	0.949	0.974

6 Conclusion

This paper has provided a comparative study on the prediction performance of several learning classifiers. This research provides a prediction model that is developed to predict autism traits. Using the Toddler Dataset July 2018, the proposed model can predict autism with 91.94%, 93.36%, 93.83%, 94.78% and 94.31% accuracy of NB, SVM, LR, RF, and k-NN algorithms respectively. Comparing all five machine learning algorithms, RF algorithm are efficient algorithms for better prediction of ASD. This result showed high performance comparing to the other existing approach of screening autism. Moreover, the proposed model can predict which attribute that can give significant results. This research also showed that Toddler dataset which having 19 attributes can be reduced to 10 attributes using feature selection method. These results can be used as references for future work on ASD screening.

Acknowledgements This work was supported by the University Malaysia Pahang with Fundamental Research Grant Scheme (FRGS): FRGS/1/2019/ICT02/UMP/02/8. Any opinions, findings, and conclusions expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agency, cooperating institution, or other individuals.

References

1. Alami M, Tahiri N, Arriaga F (2017) DIAUTIS: a fuzzy and affective multi-agent platform for the diagnosis of autism. British J Appl Sci Technol 21(4):1–28. https://doi.org/10.9734/bjast/2017/33716
2. Muthisamy TM, Jayabalan M, Rana ME (2020) Analysis of autistic spectrum disorder screening data for adolescents. Int J Curr Res Rev 12(19):23–30. https://doi.org/10.31782/IJCRR.2020.121927
3. Oma KS, Mondal P, Khan NS, Rizvi MRK, Islam MN (2019) A machine learning approach to predict autism spectrum disorder. In: 2nd International conference on electrical, computer and communication engineering, ECCE 2019, pp 7–9. https://doi.org/10.1109/ECACE.2019.8679454
4. Akyol K, Gultepe Y, Karaci A (2018) A study on autistic spectrum disorder for children based on feature selection and fuzzy. In: International congress on engineering and life science, October, pp 804–807

5. Deepa B, Jeen Marseline KS (2019) Exploration of autism spectrum disorder using classification algorithms. Procedia Comput Sci 165(2019):143–150. https://doi.org/10.1016/j.procs.2020.01.098
6. Raj S, Masood S (2020) Analysis and detection of autism spectrum disorder using machine learning techniques. Procedia Comput Sci 167(1):994–1004. https://doi.org/10.1016/j.procs.2020.03.399
7. Beaulieu-Jones B, Finlayson SG, Chivers C, Chen I, McDermott M, Kandola J, Dalca AV, Beam A, Fiterau M, Naumann T (2019) Trends and focus of machine learning applications for health research. JAMA Netw Open 2(10):e1914051.https://doi.org/10.1001/jamanetworkopen.2019.14051
8. Wingfield B, Miller S, Yogarajah P, Kerr D, Gardiner B, Seneviratne S, Samarasinghe P, Coleman S (2020) A predictive model for paediatric autism screening. Health Informatics J. https://doi.org/10.1177/1460458219887823
9. Wall DP, Kosmicki J, Deluca TF, Harstad E, Fusaro VA (2012) Use of machine learning to shorten observation-based screening and diagnosis of autism. Transl Psychiatry 2(4):e100–e108. https://doi.org/10.1038/tp.2012.10
10. Kosmicki JA, Sochat V, Duda M, Wall DP (2015) Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. Transl Psychiatry 5(2):e514–e517. https://doi.org/10.1038/tp.2015.7
11. Thabtah F (2019) Machine learning in autistic spectrum disorder behavioral research: a review and ways forward. Inform Health Soc Care 44(3):278–297. https://doi.org/10.1080/17538157.2017.1399132
12. Abdullah AA, Rijal S, Dash SR (2019) Evaluation on machine learning algorithms for classification of autism spectrum disorder (ASD). J Phys: Conf Ser 1372(1). https://doi.org/10.1088/1742-6596/1372/1/012052
13. Erkan U, Thanh DNH (2019) Autism spectrum disorder detection with machine learning methods. Curr Psychiatry Res Rev 15(4):297–308. https://doi.org/10.2174/2666082215666191111121115
14. Devika Varshini G, Chinnaiyan R (2020) Optimized machine learning classification approaches for prediction of autism spectrum disorder. Ann Autism Dev Disorders 1(1)
15. Mahadev D, Aishwarya J, Akshatha N, Anusha H, Shishira J (2020) Engagement detection with autism spectrum disorder using machine learning. Int Res J Eng Technol July:488–492. www.irjet.net
16. Thabtah F, Kamalov F, Rajab K (2018) A new computational intelligence approach to detect autistic features for autism screening. Int J Med Informatics 117:112–124
17. Schober P, Schwarte LA (2018) Correlation coefficients: appropriate use and interpretation. Anesth Analg 126(5):1763–1768. https://doi.org/10.1213/ANE.0000000000002864

The Role of Gear Grind Games in Creating a Motivating Learning Environment

W. L. WilliamCheng, P. S. JosephNg, H. C. Eaw, and K. Y. Phan

Abstract Games have played an impactful life in our current generation brimming with technological advancement. Not only has it become the main source of entertainment for people to enjoy, but it has also transitioned to becoming a learning experience for others. With our current repetitive standards and methods of teaching and learning in schools, the mind of many students has dulled their creativities as well as developing the extrinsic motivation to study at all. However, the implementation of game learning can be developed into modern-day teaching and be used as a tool for students to grasp their potentials and be motivated to learn level more using games. From the mix mode research conducted, many tertiary students would rather replace their conventional studying methods with a game learning experience. The player will be able to understand, develop listening skills and verbal communication skills towards other people to understand when it is necessary to be assertive in an environment. Moreover, this game will tackle the understanding of real-life decision-making skills that mimic real-life situations for players to adapt to. With that said, JomGear Grind Games aims to help students learn various skills such as creative thinking and interpersonal skills.

Keywords Creative problem-solving · Gamification · Game learning · Videogames · Games · Strategic problem solving · Interpersonal skill

W. L. WilliamCheng · P. S. JosephNg (✉)
Institute of Computer Science & Digital Innovation, UCSI University, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia
e-mail: josephng@ucsiuniversity.edu.my

H. C. Eaw
Faculty of Business & Management, UCSI University, UCSI Heights, 56000 Cheras, Kuala Lumpur, Malaysia
e-mail: eawhc@ucsiuniversity.edu.my

K. Y. Phan
Faculty of Information & Communication Technology, Universiti Tunku Abdul Rahman, Kampar, Perak, Malaysia
e-mail: kyphan@utar.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_13

1 The Behavioural Effects of Games on Students

The technological world has advanced so far beyond what people have comprehended. This has created a surge of innovations within the minds of many in this technological time [1]. These advancements have improved social media platforms that allow for messaging services for worldwide use to get information from various places in the world. For example, Instagram or Twitter would be a famous social media platform [2]. Adolescents and children who are still in school use this as a place of talking to new people and sharing their stories and info online. However, they have to overuse these social media platforms and have created a concern towards the media and parents for this addiction. This has become a problem to the world as these people are usually using this as a type of procrastination over their duties when the duties have been given and the responsibilities are ignored [3].

Having a creative mind that can also think strategically within a university student mind nowadays is no easy feat [4]. Skills such as having a mind to think outside the box and also have a strategic mind are sought out in workplaces and has become well-needed skills to have [5]. Within refinement of the skills, research and development will be able to have a greater impact on society and bring forth new modes of innovations from many young minds [6]. Moreover, talking with others is an essential skill to have when working with a group or with other people, with interpersonal skills coming into the mind [7]. Working together with greater efficiency is the first step in the success of a project or goal as it creates a bond and trustful relationship amongst the team [8]. Many university students come off as shy and socially awkward which deters to social and communication skills. Traumatic events or lack of human interaction and love during their early life could be a major factor in this [9]. Almost all student levels from primary to tertiary education have played games and can be hooked into getting their sole attention in understanding the concept of sand outcomes of the games [10]. It is also a fact that not only boys play video games, but girls equally do it as well which shows that there is a gender variety of video game players [11].

Self-efficacy can also be better leant and developed from video games in our current modern era as well as researched from studies [12]. It is also said that video games how developed more violent tendencies from the player due to negative habits and outcomes found within the game that tampers with the real-life behaviours of the young people that have tread a problem in the society of gaming [13, 14]. Be that as it may, positivity can be given from video games as well that helps benefit and give motivations to the player in developing interests within the game's theme and world [15].

2 Legacy and Related Works of Game Learning and Gamification

There are elements within a game which can provide a sense of learning experience and enjoyment towards gamer and it is known as gamification [16]. Conventional studying methods towards students nowadays are not going to work as they do feel enjoyment in using a textbook method to learn the outcomes of a subject and their respective skills as it may feel boring especially when technology can play a big role [17]. Because of this, university students only learn and study for a test for the sake of getting over with it and progressing in their course, this is where gamification helps in providing an intrinsic motivation to study and learn a certain course [18]. This brings forth many avenues for video games to become a platform for learning where the elements of gamification play a vital role in greater developing the motivation for the university student to learn skills [19]. User engagement can also be obtained in the form of prizes and game achievements which boosts the players desired outcome towards the goal [20]. From a study, heuristics design within the game's levels has brought forward greater refine towards the algorithm of gamification to this day [21] as shown in Table 1.

Problem Statement, Question and Objective

The certain characteristics and features of the game will emphasize the development of interpersonal, creative and strategic problem-solving skills in a game learning setting and using gamification elements to enhance the experience to the students.

The questions from Table 2 will be defined and determined from the scope of the research. The Scope of the project will mainly be university students of any field related using interpersonal, creative, and strategic problem-solving skills. The age range should be around 17–24 for typical university student ages and will take place in selected universities like Monash University, Sunway University, UCSI

Table 1 Similar gaming application comparisons table

Gaming characteristics and features	Gear grind games	Logic master	Logical test	Skills
Creativity	/	/	/	–
Design	/	/	/	/
Strategic skill	/	/	/	/
Interactive quests	/	–	–	–
Achievements	/	–	–	–
Games	/	–	/	/
scoreboards	/	–	/	–
Level building	/	/	/	/
Tasks system	/	–	–	–
Communication and dialogue	/	–	–	–

Table 2 Research questions

RQ1: Can university students grasp the understanding of creative and strategic problem solving from video games?
RQ2: Can interpersonal skills be greatly developed with the help of game learning and gamification?
RQ3: Can game learning and gamification replace the traditional studying methods in universities and provide students with intrinsic motivation to learn from games?

Table 3 Research objectives

RO1: To create a video game that emphasized the values of creative and strategic problem-solving skills.
RO2: To help university students to develop better interpersonal skills to make friends and increase teamwork via gamification.
RO3: Using game learning and gamification to intrinsically make students motivated to study traditional studying methods.

University, University Malaya, Asia pacific University and Tunku Abdul Rahman University College. The research questions will be answered by the research objectives below in Table 3.

Figure 1 shows the highlighted key points of the research which are displayed within the research model to further explain the game design aspect of the game, the software required for the game will be done by the Godot Game Engine with the usage of the Microsoft visual studio for the C# language incorporation and coding design. The main game art will be mainly done with the inspiration of pixel art games such as Games like Fire Emblem and Undertale for the retro feel of the game overall. Figma will be greatly used to test the prototyping for the game where the outline will be copied and enhance when the game undergoes user interface and world design. The game will allow the development of key skills within the research which are interpersonal skills, creative and strategic problem-solving skills within the game application to further enhance students ability to communicate and to adapt to new situations creatively while also providing an intrinsic motivation for them to learn said skills.

3 Methodology

This research will be applying the usage of mixed-mode research of surveys and interviews to gain both qualitative and quantitative data. From Fig. 2 below, a mixed-mode method of surveys and interviews will be handed off to university students that like games and are interest in gamification to obtain both qualitative and quantitative data collections. A trial will be done with 5 survey respondents and

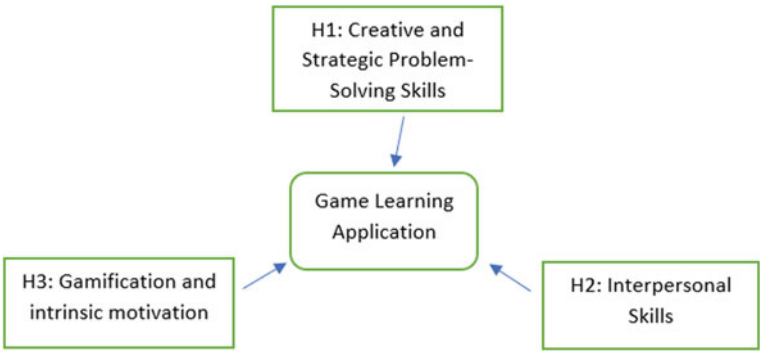


Fig. 1 Hypothesis research model

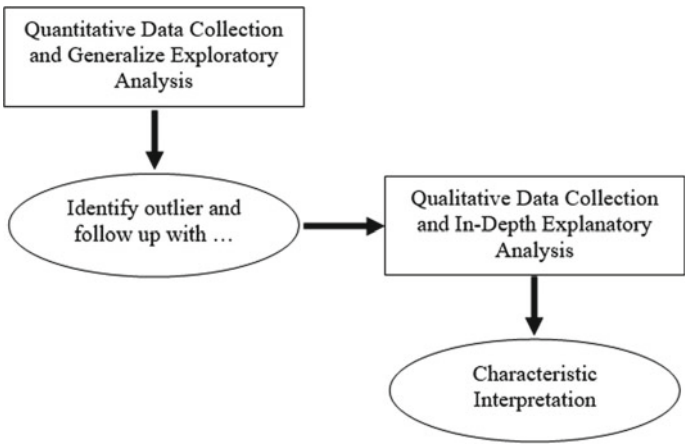


Fig. 2 Sequential design [25–37]

3 interview respondents. This is to check the efficiency of questions and check for amendments if questions are not understandable or appropriate. Random students will be selected for the research to ensure the data is not faked and the data set is qualitative.

4 Results and Findings

From the results that have been recorded from the research, the improvement of games on the creative and strategic problem solving of university students has been widely agreed on in both the preliminary and technical data collection.

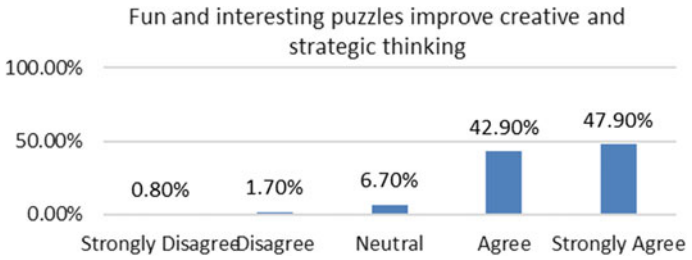


Fig. 3 Post-test results of fun puzzles improve creative and strategic thinking

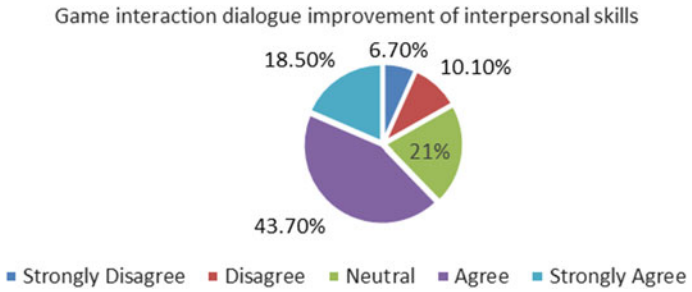


Fig. 4 Post-test results of game interaction dialogue improvement of interpersonal skills

From Fig. 3, a lot of the responses showed an agreement that games can help university students improve upon their creative and strategic problem-solving skills and more than 90% of the respondents widely accept that puzzles can help in the improvement of creative and strategic thinking. This suggests that the game application should include more interesting puzzles that can help stimulate the thinking process of university students. Meanwhile, Fig. 4 summarized the effect of game interaction dialogue on university students.

From the data given in the figure, 43.7% of respondents agreed that the dialogue shown within games and the character interaction have helped in increasing the interpersonal skills of university students due to the realism and relatability of the conversation in games. This means that the application should include non-player characters (NPC) to include interactional dialogue to the player as a sense of communication. From the interview respondents, it has shown that most respondents agree that it is known as a source of indirect communication.

Figure 5 can summarize that using games to study which leans towards game learning is better for university students as it creates more motivation for them to study compared to the usual textbook and lecture slide method. This could also motivate the lecturers at the university to use a hand on game learning method to teach students in a class so they can grasp the feel of games as an instrument of learning and also help them get more motivated to teach students as well. According to the interview results, most respondents responded with it being more

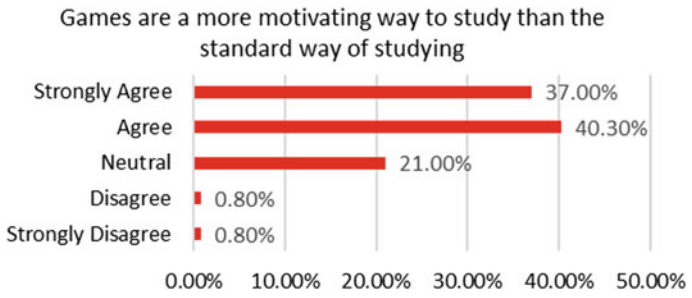


Fig. 5 Games post-test result a more motivating of studying than the standard way

fun and engaging way for students to learn while others state that games are new and not expected to the community of university students which suggests that people want more new inventive ways of studying instead of repetition.

Based on the interview and survey results of the research, most respondents agree upon the fact that most university students show a lack of creativity in out-of-the-box thinking. A total of 66% of respondents show that more people tend to follow a traditional way of studying which includes the textbook method instead of thinking of their solutions to things. This shows that the conventional way of thinking has blocked the creativity of students who only follow instructional procedures to most questions and problems. Moreover, the interview respondents showed more positive feedback towards the cost-effectiveness of game learning in universities. About 46% of respondents have agreed that having games will cut down the paper costs of printing papers in schools and promote online learning because it can save travelling as well. This would help the implementation of game learning technology in the university classroom and online platforms like Kahoot. The game app could be sold in large for universities for a low price in high quantities. The interviewers also left positive responses and feedback for remembering things that happen in games than study notes which could show that the interactivity and enjoyment of a game. Moreover, 63% of respondents agree that it helps due to the interesting features and plots of the game and it helps the brain function more due to less stress. The rest of the responses indicate that it is more visually aesthetic while books are more static in terms of memorability. This could suggest that the game application to include more interesting plot points where the learning applications are the key factors of the game at that point. Based on the responses and feedback from the data collection, the game's main menu features of the game application on a mobile and desktop mode. It includes a play option, settings option, quit option, and achievement option. Figure 6 below further illustrates the game application's main menu screen.

The achievement screen shows the allocated information of the student playing the game such as their full name, student ID, and their current lecturer in charge of the class. These points will be allocated from communication points that relate to the interpersonal skills gained in the game, creativity points that relate to the

Fig. 6 Achievement screen of the game application

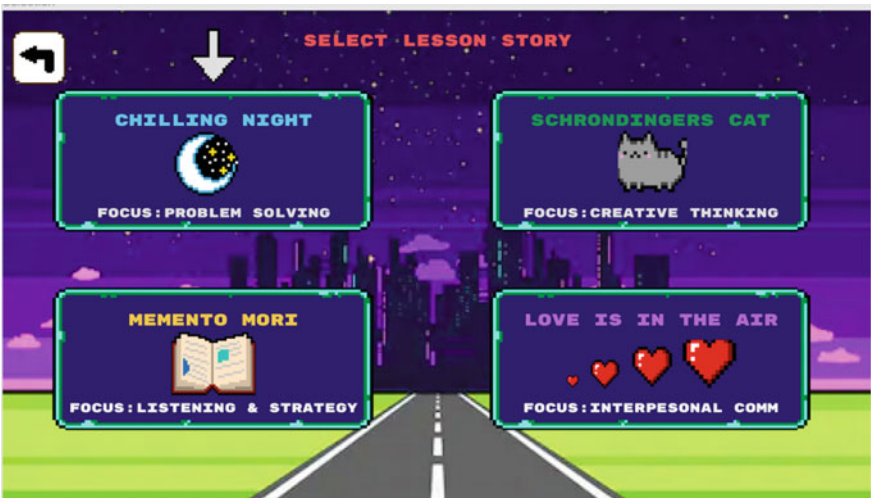


Fig. 7 Story selection of the game application

out-of-the-box thinking done in-game, and lastly the strategy points which show the tactful methods used in the game. Furthermore, the bottom part of the screen includes reward tiers that can be obtained soon which will increase the motivation factor of the students to accumulate more points to reach the hidden threshold to obtain the rewards which is a gamification factor of the game learning process.

As shown above, Fig. 7 displays the game story selection which features multiple game story modes that can be chosen by the student to play based on the focus that

they want to learn. The main focuses are shown at the bottom of each story selection so that the student will be shown a rounded-up emphasis on what they are going to get into once they select game mode. Each game mode can be played without order to give the freedom of choice to the player and the game modes can be revisited to relearn left out items and points of the story. Each story also contains its unique dialogue and to-do lists for the student to be involved in and complete. The guided storyline aims to provide a motivating environment to entice learning [38].

5 Conclusion, Limitation, and Future Work

In conclusion, JomGear Grind Games can help improve the learning habits and drive of University Students to further enhance their communication skills and out-of-the-box thinking using gamification elements that can be grasped from the game. Also, the scoreboard and rewards system help seek the encouragement of students and motivate them to try harder and more competitively as well as the storyline that lets them envelop into the game world. The limitations during the research that were found were the lack of time on how to research and the limited resources available for it. The game may be something to motivate the students to study but it is their capabilities and work that will benefit them. In the future, we seek to understand the extra benefits of the addition of AR and VR into the game for people to experience. This may benefit them even more as new technologies tend to sway the minds of people and their learning potential.

References

1. Griffy-Brown C, Earp BD, Rosas O (2018) Technology and the good society. Technol Soc Forthcoming 1–3
2. Shu MT, Yaw YW, Chih MW (2020) A study on digital games internet addiction, peer relationships, and learning attitude of senior grade of children in elementary school of Chiayi County. J. Educ Learn 9(3):2020
3. Danah H, Micahel H, Edwin C, Sona C, Miroslava C, Evgenia S, Erkko TS, Christopher HT (2018) Creativity and technology in education: an international perspective. Technol Know Learn 23:409–442
4. Leela R (2018) Cooperative learning is a constructivist strategy in tertiary education. Int J Educ Res 6(12)
5. Kang CM et al (2015) A study on integrating penetration testing into the information security framework for Malaysian higher education institutions. International symposium on mathematical science and computing research, Malaysia, pp 156–161
6. AdrianChin YK et al (2020) JomDataMining: learning behaviour affecting their academic performance, really? In: IEEE 6th international conference on engineering technologies and applied science, Malaysia
7. AdrianChin YK et al (2020) JomDataMining: academic performance and learning behavior dubious relationship. In: Int J Bus Inf Syst Forthcoming article
8. Lim LJ et al (2017) ScareDuino: smart farming with IoT. Int J Sci Eng Technol 6(6):207–210

9. Gerard JP, Cyndi B, Selcuk A, Jo AY, Molly H, John FC (2018) Creative problem-solving in small groups: the effects of creativity training on idea generation, solution creativity and leadership effectiveness. J Creat Behav 52(2):25–36
10. Matthew B (2018) Student attitudes to games-based skills development: learning from video games in higher education. Comput Human Behav 80:283–294
11. Frank B, Janet B, Michelle Q, Peter JA, Shelley BB, Josh S (2018) A clinical educator's experience using a virtual patient to teach communication and interpersonal skills. Australas J Educ Technol 34(3)
12. Rowan HB (2019) Developing teamwork skills in undergraduate science students: the academic perspective and practice. In: Proceedings of the Australian conference on science and mathematics education, Monash University, 27–29 Sept 2017, pp 137–148
13. Daphne B, Shawn CG (2019) Enhancing attentional control lessons from action video games. Neuron 104:147–163
14. Fonseca D, Villagrasa S, Navarro I, Sánchez A (2017) Urban gamification in architecture education. Adv Intell Syst Comput 571(1):335–341
15. Polo-Pena AI, Frias-Jamilena DM, Fernandez-Ruano ML (2020) Influence of gamification on perceived self-efficacy: gender and age moderator effect. Int J Sports Mark Sponsorship
16. Kühn S, Kugler D, Schmalen K et al (2019) Does playing violent video games cause aggression? A longitudinal intervention study. Mol Psychiatry 24:1220–1234
17. Lei W, Kunter G, Ramesh S, Joseph P (2020) Impact of gamification on perceptions of word-of-mouth consumers. Manag Inf Syst 44(4):1–33
18. Devin JM, Marina M, Jessica M, Nancy LH (2018) Exploring the pull and push underlying problem video game use: a self-determination theory approach. Personality Individ Differ 135:176–181
19. Rapp A, Hopfgartner F, Hamari J et al (2018) Strengthening gamification studies: current trends and future opportunities of gamification research. Int J Hum Comput Stud 126:1–6
20. Lin C-J, Hwang G-J, Fu Q-K, Chen J-F (2018) A flipped contextual game-based learning approach to enhancing EFL students' english business writing performance and reflective behaviors. Educ Technol Soc 21(3):117–131
21. Waweru BW et al (2020) Gamesy: how videogames serve as a better replacement for school? In: IEEE student conference on research and development, pp 10–15
22. Toda AM, Klock ACT, Oliveira W et al (2019) Analysing gamification elements in educational environments using an existing Gamification taxonomy. Smart Learn Environ 6 (16):18–32
23. Mohd Hishamuddin AR, Ismail YP, Noor Anida ZMN, Nor Syazwani MS (2018) Gamification elements and their impacts on teaching and learning—a review. Int J Multimedia Appl (IJMA) 10(6)
24. Knutas A, van Roy R, Hynninen T et al. A process for designing algorithm-based personalized gamification. Multimed Tools Appl 78:13593–13612
25. JosephNg PS, Loh YF, Eaw HC (2020) Grid computing for MSE during volatile economy. In: International conference on control, automation and systems, IEEE Explore, Busan, Korea, pp 709–714
26. JosephNg PS, Kang CM, Mahmood AK, Wong SW, Phan KY, Saw SH, Lim JT (2016) EaaS: available yet hidden infrastructure inside MSE. In: 5th International conference on network, communication, and computing, ACM international conference proceeding series, Kyoto, Japa*n*, pp 17–20
27. JosephNg PS, Kang CM, Choo PY, Wong SW, Phan KY, Lim EH (2015) Beyond cloud infrastructure services in medium-size manufacturing. In: International symposium on mathematical sciences and computing research, IEEE Explore, Ipoh, Malaysia, pp 150–155
28. JosephNg PS, Choo PY, Wong SW, Phan KY, Lim EH (2012) Malaysia SME ICT during economic turbulence. In: International conference on information and computer network, Singapore, pp 67–71

29. Ng Poh Soon J, Yin CP, Wan WS, Nazmudeen MSH (2011) Energizing ICT infrastructure for Malaysia SME during economic turbulence. In: Student conference on research and development, IEEE Explore, Cyberjaya, Malaysia, pp 328–322
30. JosephNg PS, Eaw HC (2021) Making financial sense from EaaS for MSE during economic uncertainty. Adv Intell Syst Comput Forthcoming
31. JosephNg PS, Eaw HC (2021) Still technology acceptance model? Reborn: exostructure as a service model. In: International journal of business information systems, Forthcoming
32. JosephNg PS (2019) EaaS infrastructure disruptor for MSE. Int J Bus Inf Syst 30(3):373–385
33. JosephNg PS (2018) EaaS optimization: available yet hidden information technology infrastructure inside medium-size enterprises. J Technol Forecast Soc Change 132(July): 165–173
34. JosephNg PS, Kang CM, Mahmood AK, Choo PY, Wong SW, Phan KY, Lim EH (2016) Exostructure services for infrastructure resources optimization. J Telecommun Electron Comput Eng 8(4):65–69
35. JosephNg PS, Kang CM (2016) Beyond barebone cloud infrastructure services: stumbling competitiveness during economic turbulence. J Sci Technol 24(1):101–121
36. Joseph NPS, Mahmood AK, Choo PY, Wong SW, Phan KY, Lim EH (2015) Barebone cloud IaaS: revitalization disruptive technology. Int J Bus Inf Syst 18(1):107–126
37. Joseph NPS, Mahmood AK, Choo PY, Wong SW, Phan KY, Lim EH (2014) IaaS cloud optimization during economic turbulence for malaysia small and medium enterprise. Int J Bus Inf Syst 16(2):196–208
38. Mitchell R, Schuster L, Jin HS (2018) Gamification and the impact of extrinsic motivation on needs satisfaction: defining gamification. J Bus Res 3(2):323–330

Sensitivity Analysis of Intelligent Active Force Control Applied to a Quadrotor System

Sherif I. Abdelmaksoud, Musa Mailah, and Ayman M. Abdallah

Abstract This paper introduces a hybrid controller to be employed for stabilizing a quadrotor, as an example of rotor unmanned aerial vehicle (UAV) systems, and efficiently repelling the applied perturbations during trajectory tracking in a complex environment via a simulation study. In this study, *Newton-Euler's* method was used to find the equations of motion for the dynamic model of the quadrotor system taking into account the effects of aerodynamic, gyroscopic, perturbation, and friction. The proposed control structure was comprised of a proportional-integral-derivative (PID) control scheme and an innovative control technique known as active force control (AFC). The AFC was tuned intelligently using artificial intelligence (AI)-based approach, namely, iterative learning algorithm (ILA), to be defined as intelligent active force control (IAFC), and the proposed strategy was identified as (PID-ILAFC) scheme. To evaluate the feasibility of the proposed control scheme, a sinusoidal wave disturbance was introduced as an example of external perturbation. In this study, a sensitivity analysis was also performed considering the variance of the estimated inertia value, model uncertainty, and the AFC output signal percentage to achieve the best possible performance. The outcomes show the efficacy of the IAFC-based strategy in expelling off the applied perturbations and uncertainties in comparison with the PID controller. The results also indicate the importance of the accurate selection of the estimated inertia value to enhance the effectiveness of the AFC approach and the significance of using the full output of the AFC signal to obtain the best performance.

Keywords Active force control · Artificial intelligence · PID controller · Quadrotor · UAVs · Iterative learning · Trajectory tracking · Sensitivity analysis

S. I. Abdelmaksoud (✉) · M. Mailah
School of Mechanical Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Malaysia
e-mail: iasherif@graduate.utm.my

A. M. Abdallah
Aerospace Engineering Department, King Fahd University of Petroleum and Minerals, Dhahran 31261, Saudi Arabia
e-mail: aymanma@kfupm.edu.sa

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_14

1 Introduction

Quadrotors have received preferred research attentions, for indoor and outdoor applications, over the past few decades, compared to other UAV models, because they are characterized by high maneuverability, vertical take-off and landing, ability to hover at a certain point, low servicing time and cost, lightweight. Thus, they have been employed in several fields such as inspection processes, shipping and delivery, precision agriculture, weather forecasting, entertainment, disaster management, and military sections [1, 17].

With the continuous improvement of the quadrotors industry, they face several challenges, such as bounded operating speed, battery life, durability, and limited flight range, and restriction of electrical motors [12]. Among several impediments, following the anticipated trajectories while eliminating different shapes of external perturbations is deemed necessary for the quadrotors since they are susceptible to external disturbances, noises, and uncertainties during trajectory tracking. This may, in turn, affect their performances and cause undesirable movements leading sometimes to failure in the whole dynamic systems.

2 Literature Review/Related Work

For disturbance rejection, an active disturbance rejection control (ADRC) strategy has been planned to successfully cancel the external disturbances while following certain paths [6, 21]. Dong et al. [10] suggested, experimentally and analytically, a flight control unit employing an adjusted ADRC consisted of a PD controller, a time-optimized tracking differentiator, and an extended state observer for a quadrotor subjected to input delays and external turbulences. While Alkamachi and Erçelebi [8] proposed a H_∞ controller to study its ability in rejecting model uncertainties and disturbances. However, the ADRC structure is complex and needs to tune a large number of control parameters. Moreover, it is not quite suitable for operations with large time delays. Furthermore, Abdelmaksoud et al. [5] proposed a self-regulating fuzzy PID (SR-FPID) controller to improve system stability and disturbance rejection capability.

One of the innovative methods is an active force control (AFC) technique. In the early 1980s, it was first proposed by Hewitt and Birdies for a rigid robot arm [14]. The AFC method is based on finding the appropriate and accurate value of the inertia or mass parameter of the model through an application of the trial and error (TEM) method [16] or any artificial intelligence (AI) method [2, 3, 15], and the measured force/torque and acceleration signals of the model. One of the distinguishing characteristics of the AFC method is its ability to integrate with classic, modern, or intelligent control methods allowing it to be used in several applications. Research works are reported in [2, 16] that utilized the AFC-based control unit for controlling the quadrotor UAVs as results proved its effectiveness. Thus, the AFC is

Table 1 Comparison between the ADRC and AFC strategies [18, 19]

	Benefits	Drawbacks
ADRC	• Accomplish superior control performance • Efficient in eliminating applied perturbations and uncertainties	• The control scheme is complicated and needs to adjust many control parameters • Difficult to be implemented in real-time • Not completely suitable for operations with large time delays
AFC	• Ease of application in practice • Adjusting the estimated inertia is the only control parameter • Effective in eliminating different types of external or internal perturbations and uncertainties • Ability to merge with classical, modern, or smart control systems	• Actuator saturation • Obtaining accurate measurements of force/torque and acceleration signals

distinguished by ease of implementation in practice. It is also efficient in rejecting various forms of external or internal disturbances, and uncertainties. Moreover, it has the ability to merge with classical, modern, or intelligent control strategies. The main differences between the ADRC and AFC strategies were listed in Table 1.

The impact of this article is in the implementation of a novel control scheme based on the AFC approach tuned intelligently to settle the quadrotor structure and effectively reject the undesired disturbances. Additionally, the other contribution is to investigate the sensitivity analysis to obtain the best possible performance of the proposed AFC-based control strategy taking into account the variance in the estimated inertia value, model uncertainty, and the AFC signal output percentage.

The article's organization is set out as follows: Sect. 2 explains the kinematics and dynamics of the quadcopter considering specific assumptions, and the proposed AFC with iterative learning (IL) implementation. The simulated results of the system responses are presented in Sect. 3. While the sensitivity analysis is exhibited in Sect. 4. Finally, the paper's conclusion is shown in Sect. 5.

3 Methodology

In this section, the mathematical model of the quadcopter model was derived based on certain considerations. In addition, an intelligent hybrid control system was proposed to adjust the behavior of the quadcopter in the existence of external perturbations and uncertainties. Finally, simulation procedures to implement the suggested control systems for the derived model were discussed.

3.1 Quadrotor Modeling

In this part, the dynamics of the quadcopter system were discussed including aerodynamic, perturbation, gyroscopic, and drag effects.

To derive the quadrotor system, certain assumptions need to be considered [7]:

- The axes and center of gravity of the quadcopter system correspond with the axes and origin of the body-fixed frame.
- The quadcopter system is inflexible and symmetrical.
- The propellers of the rotor are rigid.

The quadcopter has two subsystems; the translational subsystem that defines its position (x, y, z) and the rotational subsystem which expresses its orientations (ϕ, θ, ψ), thus it is a six degrees of freedom (DOF) system. The quadcopter system has six DOF movements with four control actions leading it to be considered as an underactuated model.

The nonlinear translation movements of the aerial system are given as:

$$\sum F^{\mathrm{B}} = m\dot{v}^{B} + \omega^{\mathrm{B}} \times \left(mv^{\mathrm{B}}\right) \tag{1}$$

where ω is the angular velocity vector.

The angular momentum of a body with inertia matrix J in the rotational subsystem can be expressed as:

$$\sum M^{\mathrm{B}} = J\dot{\omega}^{\mathrm{B}} + \omega^{\mathrm{B}} \times \left(J\omega^{\mathrm{B}}\right) \tag{2}$$

Therefore, the derived translational and rotational equations of motion considering the gyroscopic, aerodynamic, drag, and disturbance effects can be expressed as [1, 13]:

$$\ddot{x} = \frac{U_1}{m}(\cos\phi\ \sin\theta\ \cos\psi + \sin\phi\ \sin\psi) - k_1\dot{x} + d_1 \tag{3}$$

$$\ddot{y} = \frac{U_1}{m}(\cos\phi\ \sin\theta\ \sin\psi - \sin\phi\ \cos\psi) - k_2\dot{y} + d_2 \tag{4}$$

$$\ddot{z} = \frac{U_1}{m}(\cos\phi\ \cos\theta) - \mathrm{g} - k_3\dot{z} + d_3 \tag{5}$$

$$\ddot{\phi} = M_{\mathrm{dp}} + \frac{lU_2}{I_{\mathrm{xx}}} - \frac{\dot{\theta}J_{\mathrm{r}}w_{\mathrm{r}}}{I_{\mathrm{xx}}} + \dot{\psi}\dot{\theta}\left(\frac{I_{\mathrm{yy}} - I_{\mathrm{zz}}}{I_{\mathrm{xx}}}\right) - k_4\dot{\phi} \tag{6}$$

$$\ddot{\theta} = M_{\mathrm{dq}} + \frac{lU_3}{I_{\mathrm{yy}}} + \frac{\dot{\phi}J_{\mathrm{r}}w_{\mathrm{r}}}{I_{\mathrm{yy}}} + \dot{\psi}\dot{\phi}\left(\frac{I_{\mathrm{zz}} - I_{\mathrm{xx}}}{I_{\mathrm{yy}}}\right) - k_5\dot{\theta} \tag{7}$$

$$\ddot{\psi} = M_{\mathrm{dr}} + \frac{U_4}{I_{zz}} + \dot{\phi}\dot{\theta}\left(\frac{I_{\mathrm{xx}} - I_{\mathrm{yy}}}{I_{zz}}\right) - k_6\dot{\psi} \tag{8}$$

where

U_1, U_2, U_3, and U_4: altitude, rolling, pitching, and yawing control actions, respectively
g: gravitational acceleration
m: mass of the quadcopter
l: distance from the center of mass to each motor
k_1, k_2, and k_3: aerodynamic translational coefficients
k_4, k_5, and k_6: aerodynamic rotational coefficients
d_1, d_2, and d_3: input disturbances
J_{r}: rotors' inertia
M_{dp}, M_{dq}, and M_{dr}: random disturbance moment
$I_{\mathrm{xx}}, I_{\mathrm{yy}}$, and I_{zz}: moment of inertia through the principal axes in the body frame
w_r: rotor relative speed

The translational subsystem is clearly underactuated, as there are three DOF with one control action, while the rotational subsystem is fully actuated.

3.2 *Proposed Active Force Control with Iterative Learning Control (ILAFC)*

In this study, the proposed hybrid control scheme was based on a designed PID controller cascaded with the AFC strategy. To successfully employ the AFC-based control strategy, it is essential to select the precise value of the estimated inertia. Thus, the IL-based AI method was employed to automatically determine the estimated inertia value in addition to the AFC-based scheme adjusted with a crude approximation technique. The iterative learning control (ILC) is an intelligent control approach that intelligently works to improve system performance using the outputs of previous iterations [20]. In addition, it gives a solution to a system subjected to different types of inputs.

Because of the similarity of the mathematical formula to a standard PID controller, iterative learning algorithm (ILA) can be represented as P, PD, or PID-type ILC [9]. In this research, a PD-type ILC was proposed and incorporated into the ILAFC system to automatically provide the proper value for the estimated inertia matrix (IM) where γ and β are learning constants. It can be expressed, as follows:

$$\mathrm{IM}_{\mathrm{k}+1} = \mathrm{IM}_{\mathrm{k}} + \left(\gamma + \beta\frac{d}{dt}\right)e_{\mathrm{k}}(t) \tag{9}$$

Table 2 Quadrotor system parameters [11, 22]

Description	Unit	Value
Quadrotor mass, m	kg	0.65
Distance from the center of mass to each motor, ℓ	m	0.23
Moment of inertia about the x-axis, I_{xx}	kgm^2	7.5E-3
Moment of inertia about the y-axis, I_{yy}	kgm^2	7.5E-3
Moment of inertia about the z-axis, I_{zz}	kgm^2	1.3E-2
Gravity, g	m/s^2	9.8
Thrust force coefficient, K_F	Ns^2	3.13E-5
Moment coefficient, K_M	Nms^2	7.5E-7
Moment of Inertia of the rotor, J_r	kgm^2	6E-5
Translational drag force coefficients, K_1, K_2, K_3	Ns/m	0.01
Aerodynamic friction coefficients, K_4, K_5, K_6	Ns/rad	0.012
Attitude's saturation	$20^\circ \geq \phi, \theta, \psi \geq -20^\circ$	

where

IM_{k+1}: new value of the estimated inertia

IM_k : old value of the estimated inertia

3.3 Simulation

The simulation environment was executed using MATLAB/Simulink software where the designed PID gains were tuned heuristically. The PID, PID-AFC, and PID-ILAFC were implemented and simulated to analyze the feasibility of the planned control strategies. The quadcopter parameters are tabulated in Table 2. To evaluate the effectiveness of the control methods, an external perturbation was added to the quadcopter model in the form of the sine wave. It is noteworthy that all outcomes were simulated with respect to time (time-domain) using the fourth/fifth order of the ode with a variable step solver and a relative tolerance of 0.001.

4 Results and Discussion

The time responses related to x (forward motion), y (sideward motion), z (upward motion), and *Psi* (yawing motion) of the derived quadrotor system, using the PID, PID-AFC, and PID-ILAFC systems, in the existence of the sinusoidal wave perturbation, are shown in Fig. 1.

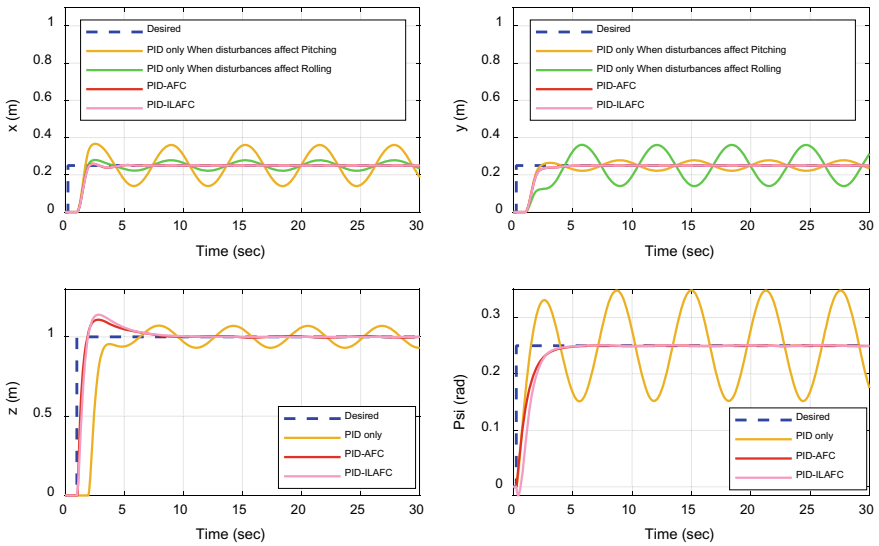


Fig. 1 Time responses of the PID, PID-AFC, and PID-ILAFC control schemes in the existence of sine wave perturbation

Based on the observed results, it can be concluded that the AFC strategy-based suggested control systems show improvement in performance in comparison to the standard PID controller. Furthermore, the proposed PID-AFC and PID-ILAFC display astonishing performance in suppressing the oscillations, vibrational levels, and applied perturbations. Moreover, the findings indicate the effectiveness and superiority of the PID-ILAFC method in enhancing performance, settling the aerial system, and eliminating the external perturbations when compared to the other control strategies. The efficacy of the proposed ILAFC-based strategy stems from its reliance on the ILA method which is characterized by automatic tuning, efficiency, and simplicity. Furthermore, the proposed ILAFC-based scheme has the advantage of being able to set the control parameters automatically and on-line, based on the specified loading and operating situations. This feature is essential for experimental work at variance to other off-line control strategies [4].

5 Sensitivity Analysis

In this section, three different cases were conducted for the sensitivity analysis, namely, the variation in the estimated inertia value, in the payload mass, and in the AFC output signal percentage. For all simulated cases, the PID-AFC scheme was utilized wherein it was assumed that the disturbance was applied only in the *Z*-direction (upward motion).

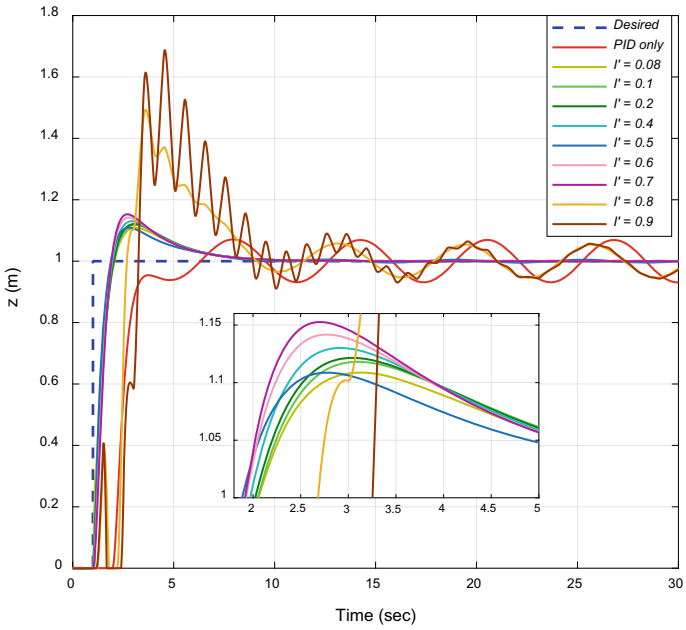


Fig. 2 System responses with a variation of estimated inertia values

5.1 Variation in the Estimated Inertia Value

In this part, different values of the estimated inertia (I′) were examined to assess the effect of this variance on the efficacy of the AFC-based control scheme. The system responses with different values of the estimated inertia are shown in Fig. 2. The results imply that when the I' increases, the effectiveness of the AFC becomes better though further increases lead to oscillatory responses due to the direct relationship between the I′ and system stability [2]. Moreover, based on the results, the best possible performance was achieved when I′ was approximately between 0.1 and 0.7.

5.2 Variation in the Payload Mass

In this section, different values of a payload mass were added to the quadrotor model (m) to verify the efficiency of the AFC-based method in the presence of model uncertainties and disturbances. The system responses for different values are revealed in Fig. 3. The results indicate that when the payload mass values increase, the AFC-based controller can settle the quadcopter while rejecting the external perturbations and uncertainties effectively.

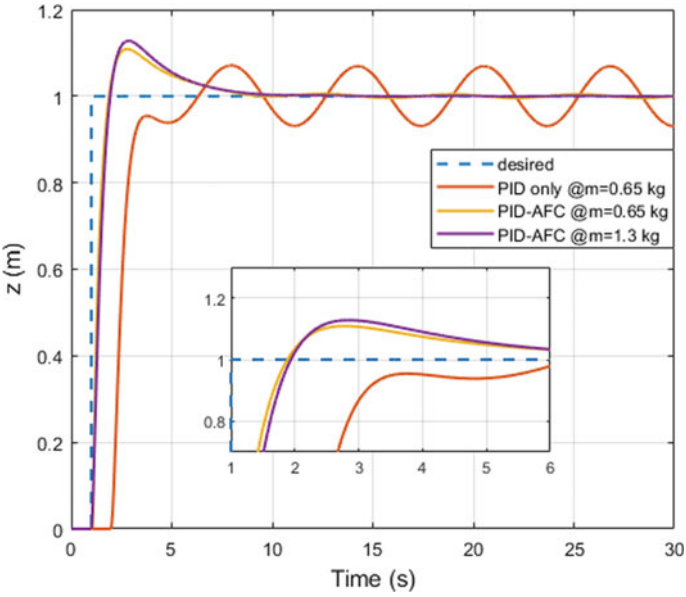


Fig. 3 System responses with a variation of the payload mass values

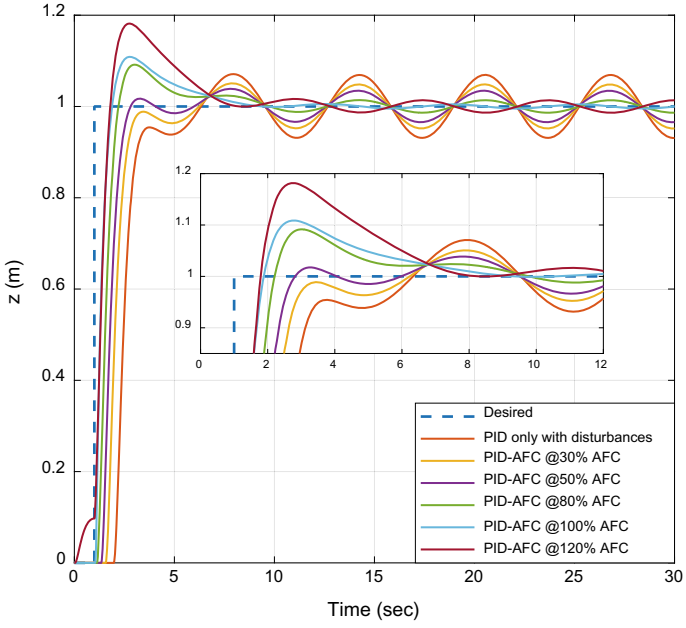


Fig. 4 System responses with a variation of the AFC percentages

5.3 *Variation in the AFC Signal Output Percentage*

The impact of utilizing the appropriate output of the AFC control signal on the dynamic system performance was investigated. The output of the AFC signal was considered 100 percent before proceeding to the control system.

For example, when the output of the AFC signal is used and passed as is, this is considered a 100% utilization of the AFC signal. Four cases, namely, 30, 50, 80, and 120% of the AFC control signal were used and compared with the total output (100%) of the AFC signal counterpart, as shown in Fig. 4. Based on the responses of the system, it can be concluded that the use of 100% of the AFC signal is the best selection and highly recommended to be implemented in settling the dynamic model and eliminating the applied perturbations.

6 Conclusion

The applied perturbations and gusty atmospheres are impediments that affect the performance of the quadcopter during following certain paths. Moreover, the quadcopter is an underactuated, highly non-linear, and coupling dynamic model, and it is also sensitive to noises, disturbances, and uncertainties. Thus, the need for a control strategy capable of stabilizing the aerial model and rejecting several forms of external disturbances is essential. In this work, a novel intelligent control structure has been proposed based on the AFC strategy. The proposed PID-AFC and PID-ILAFC methods have been effectively designed and implemented where the outcomes reveal the efficacy of the recommended schemes in stabilizing the quadrotor model and rejecting the introduced disturbance in comparison to the standard PID control system. Furthermore, different cases of the sensitivity analysis have been conducted showing the importance of the appropriate and accurate selection of the estimated inertia value, the efficacy of the AFC technique in the presence of the model uncertainties, and the importance of using the full output of the AFC signal to obtain the best possible performance. Future work may concentrate on applying the suggested PID-ILAFC approach experimentally on a real quadcopter model for validation. Also, further study should be conducted to investigate several forms of disturbances and various operating and loading circumstances.

References

1. Abdelmaksoud SI, Mailah M, Abdallah AM (2020) Control strategies and novel techniques for autonomous rotorcraft unmanned aerial vehicles: a review. IEEE Access. 8:195142–195169
2. Abdelmaksoud SI, Mailah M, Abdallah AM (2020) Robust intelligent self-tuning active force control of a quadrotor with improved body jerk performance. IEEE Access. 8:150037–150050

3. Abdelmaksoud SI, Mailah M, Abdallah AM (2020) Disturbance rejection for a quadrotor using robust active force control with genetic algorithm. Int J Model Ident Control 36(3):200. https://doi.org/10.1504/IJMIC.2020.116915
4. Abdelmaksoud SI, Mailah M, Abdallah AM (2021) Practical real-time implementation of a disturbance rejection control scheme for a twin-rotor helicopter system using intelligent active force control. IEEE Access. 9:4886–4901
5. Abdelmaksoud SI, Mailah M, Abdallah AM (2021) Improving disturbance rejection capability for a quadcopter UAV system using self-regulating fuzzy PID controller. In: 2020 international conference on computer, control, electrical, and electronics engineering (ICCCEEE), pp 1–6. https://doi.org/10.1109/ICCCEEE49695.2021.9429661
6. Aboudonia A, El-Badawy A, Rashad R (2017) Active anti-disturbance control of a quadrotor unmanned aerial vehicle using the command-filtering backstepping approach. Nonlinear Dyn 90(1):581–597
7. Alkamachi A, Erçelebi E (2017) Modelling and genetic algorithm based-PID control of h-shaped racing quadcopter. Arab J Sci Eng 42(7):2777–2786
8. Alkamachi A, Erçelebi E (2018) H∞ control of an overactuated tilt rotors quadcopter. J Cent South Univ 25(3):586–599
9. Arimoto S, Kawamura S, Miyazaki F (1986) Convergence, stability and robustness of learning control schemes for robot manipulators. In: Proceedings of the international symposium on robot manipulators on recent trends in robotics: modeling, control and education, pp 307–316. Elsevier North-Holland, Inc., New York, NY, USA
10. Dong W, Gu G-Y, Zhu X, Ding H (2016) A high-performance flight control approach for quadrotors using a modified active disturbance rejection technique. Robot Auton Syst 83:177–187
11. Elkholy H, Talla MN (2014) Dynamic modeling and control of a Quadrotor using linear and nonlinear approaches. American University in Cairo
12. Emran BJ, Najjaran H (2018) A review of quadrotor: an underactuated mechanical system. Annu Rev Control 46:165–180
13. Hasseni S-E-I, Abdou L (2018) Decentralized pid control by using ga optimization applied to a quadrotor. J Autom Mobile Robot Intell Syst 12(2):33–44
14. Hewit JR, Burdess JS (1981) Fast dynamic decoupled control for robotics, using active force control. Mech Mach Theory 16(5):535–542
15. Mailah M (1998) Intelligent active force control of a Rigid robot arm using neural network and Iterative learning algorithms. University of Dundee
16. Omar M, Mailah M, Abdelmaksoud SI (2017) Robust active force control of a quadcopter. Jurnal Mekanikal. 40(2):12–22
17. Saeed AS, Younes AB, Cai C, Cai G (2018) A survey of hybrid Unmanned Aerial Vehicles. Prog Aerosp Sci 98:91–105
18. Tan W, Fu C (2015) Analysis of active disturbance rejection control for processes with time delay. In: 2015 American control conference (ACC), pp 3962–3967. IEEE, Chicago, IL, USA
19. Wang F, Wang R-J, Liu E-H (2019) Analysis and tuning for active disturbance rejection control. Math Probl Eng 2019:1–11
20. Xu J-X, Tan Y (2003) Linear and nonlinear iterative learning control. Springer, Berlin, Heidelberg
21. Zhang Y, Chen Z, Zhang X, Sun Q, Sun M (2018) A novel control scheme for quadrotor UAV based upon active disturbance rejection control. Aerosp Sci Technol 79:601–609
22. Zhou Z, Wang H, Hu Z, Wang Y, Wang H (2019) A multi-time-scale finite time controller for the quadrotor UAVs with uncertainties. J Intell Robot Syst 94:521–533

Scalable Smartification of Commercial Buildings HVAC Systems Using the Internet of Things and Machine Learning

Mohammad Najah Mahdi, Taofiq Adeola Bakare, Abdul Rahim Ahmad, Adamu Muhammad Buhari, and Khalid Sheikhidris Mohamed

Abstract Most of the commercial buildings in Malaysia are still equipped with the legacy control for their Heat ventilation and air conditioning system (HVAC), which several studies claimed to have contributed to energy consumption in buildings. A significant amount of this energy is consumed by the building Heat Ventilation and Air Conditioning units. This is mostly due to the lack of smart and remote functionalities in the legacy HVAC systems to control the chillers and the Air handling units. This massive energy consumption is an antithesis to what governments all over the world are aiming for. However, scalability and deployment of low-cost resource-limited hardware embedded with control algorithms used to save energy in commercial building's Heat Ventilation and Air Conditioning (HVAC) units is a difficult engineering task. But the unprecedented advancement and perverseness of information technology services over the past two decades has led to an ever more connected world. This project will leverage the concept of the Internet of Energy to make the systems smarter and more decentralized for flexible energy usage. Modern-day devices are increasingly linked to the internet, creating what is now referred to as the internet of things (IoT). The IoT paradigm has provided technologists with the ability to remotely control devices, and with the recent progress in Machine learning (ML) and Artificial Intelligence (AI), devices are trained to make smart decisions that can independently influence human to machine interactions.

M. N. Mahdi (✉)
Informatics and Computing in Energy, Universiti Tenaga Nasional, Selangor, Malaysia
e-mail: mehdi.najah@uniten.edu.my

T. A. Bakare · A. M. Buhari · K. S. Mohamed
Faculty of Engineering, Multimedia University, Cyberjaya, Malaysia

A. R. Ahmad
Systems and Network Department, College of Computing and Informatics (CCI), Universiti Tenaga Nasional, Selangor, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_15

Keywords Artificial intelligence · Energy consumption · Internet of energy · Internet of things · Machine learning · Heat ventilation and air conditioning

1 Introduction

The estimated global energy consumption by the building sector both commercial and residential buildings is about 20% [1]. Couple this with accelerated population and economic growth, and we have a projected 1.3% increment in building's energy consumption from 2018 to 2050 [2]. In addition to the massive energy consumption, the building sector accounts for two-thirds of halo-carbon, one-third of greenhouse gases, and 25–33% of black carbon emissions [3]. The energy consumption of commercial buildings in Malaysia rose from 91,539 to 108,732 GWH in 2011; while the overall energy consumption of the country is estimated at 116 Million tons of oil equivalent in 2020 [4]. Buildings in the United States alone account for 40% of the energy consumption, of which half of this is used up by Heat Ventilation and Air Conditioning (HVAC) systems [5]. Also, the growing energy costs, increase in restrictive environmental regulations, and request by consumers for better green and sustainable services are factors driving the calls for efficient energy management [6].

In light of this need for better energy management, a wide variety of Building Energy Analysis (BEA) modelling approaches have been developed by different scholars since the early 1990s [7]. Curram and Mingers [8] hypothesized that researchers can leverage current technologies to reduce electricity consumption in buildings by as much as 30–80%. This is a promising estimate but to achieve the energy reduction goal, an efficient and cost-effective system is needed. This research focuses on the following problem statements:

1. Legacy HVAC systems lack communication and smart functions needed for efficient energy-saving operations
2. The relationship between several factors affecting the efficient operation of HVAC systems is complex and non-linear

The rest of the paper is organized as follows: Sect. 2 presents a research background. Section 3 introduces the research methodology details on the project description, challenges, and concerns, and the preliminary simulations are described in the subsequent sections; finally, conclusion and future work in Sect. 4.

2 Research Background

Many approaches proposed in the literature to improve the energy efficiency of the HVAC system use a simplified thermal dynamics model to predict the temperature evolution of buildings [9–13]. In Barrett and Linder [14] the authors designed and

developed a nonlinear model of the building's HVAC system including the chillers, air handling units, and different temperature zones: the solution to this problem is a token-based Model Predictive Control (MPC) scheme to improve energy efficiency in the building.

Li and Xia [15] used the equivalent thermal parameter model to simulate air conditioners for Demand Response Programs (DR). And Jagarajan et al. [16] proposed a simplified thermal model in the stochastic state-space and performance maps for predicting the coupled thermal response of a cooling unit. While Husin et al. [17] implemented an aggregated model of a house cooling unit to establish the relationship between the external environment, indoor temperature, and the power consumption of the cooling system.

Borowski et al. [18] proposed a dynamic control strategy to participate in demand response schemes by using a simple model of an air conditioning unit which was then further scaled to include many such units. The accuracy and performance of these models rely heavily on the thermal dynamic models and mathematics behind it. However, the occupant's comfort level and the temperature of the building are affected by several random variables that are very difficult to model accurately. Therefore, "it is often intractable to develop a building dynamics model that is both accurate and efficient enough for effective run time of HVAC control" [19]. This nonlinear and highly complex relationship between the environment and energy consumption of the HVAC system has encouraged researchers to actively explore the use of real-time data-driven methods with the use of Reinforcement machine learning models to approach this problem [20].

Feng et al. [21] used classical Q-learning to optimize the reward function to control the HVAC system but this is not suitable for cases with large state space. The heat transfer process was modelled in Ascione et al. [22] but the authors only evaluated the solution with a single-zone building. However, in 2015 a Deep Reinforcement Learning (DRL) algorithm defeated the world champion in the game of Go. This model shows great versatility in solving complex and nonlinear problems.

Importance of Internet of things and Machine learning (1) Internet of things can be used to transform legacy BAS into a more smart and efficient system. (2) Machine learning algorithms are effective in solving non-linear, complex problems. (3) The low power consumption, versatile, and open network architecture of the IoT devices can transform traditional BAS control architecture to a more decentralized and smart system to conserve energy in buildings.

3 Methodology

The solutions proposed in the literature are classified into three; white box, black box, and grey box [23–27]. The white box approach involves the design of the physics model of buildings. The model helps to form the problem, understand the improvements needed, and monitor the performance of the solution. The downside

to this approach is that the relationship between the thermal variables of a building is complex and non-linear. Furthermore, this solution cannot be scaled because buildings have different characteristics [28–31].

The black box approach investigates the building-energy-related model without understanding its internal relationships. While the grey box approach or hybrid approach incorporates the properties of both the white and black box approach. These models were designed to eliminate the limitations that exist in either of them. Fazenda et al. [32] created the model of a building using Energy Plus and fed the variables affecting the comfort and energy consumption of the building into a Reinforcement machine learning model. The model interacts with these variables to learn the relationships between them to create the best energy-saving policies. However, the centralized nature of HVAC systems is the main obstruction to accurate monitoring and control. This is mostly due to the lack of smart and remote functionalities in the legacy HVAC systems to control the chillers and the Air handling units [33]; Which usually leads to massive leakage in electricity consumption. But the advances in the field of information technology and engineering has brought about the paradigm shift to the Internet of Things (IoT).

The integration of the Internet for electrical energy transactions has contributed positively to the global energy market in many ways. Internet of Energy (IoE) seeks to build a self-managed smart electrical infrastructure that minimizes energy waste through the use of Building Automation Systems (BAS) [34]. IoE has brought about a revolution in power generation, storage, and transmission through the incorporation of sensors, the Cloud, smart grids, clean energy infrastructure, and Advanced Metering Infrastructure (AMI) (see Fig. 1). Furthermore, the massive but recent proliferation of distributed renewable energy sources is shifting the focus from the traditional top to bottom generation and distribution of electricity to a more decentralized system. However, with new technology comes new challenges. The most prevalent issues facing the IoE include Demand-side management, P2P sharing, Big data, scalability and security.

Accumulation of large energy data leads to the challenge of how to manage these data to get meaningful and actionable information. This presents a huge opportunity for data-driven approaches over traditional physics-based methods. This project will leverage the concept of the Internet of Energy and Machine learning to make the systems smarter and more decentralized for flexible energy usage. The projected outcome of this study is to produce:

(1) Multi-zone control devices for data collection and actuation: To enable decentralized control of the HVAC system, the building is compartmentalized into zones depending on its weather condition. These data are collected by the sensors, processed and acted upon by the actuators retrofitted into the rooms.
(2) Machine learning algorithms for data processing and prediction: The Machine learning algorithms are used to analyze the big data, predict the energy consumption, and the suitable temperature set-points for the zones.

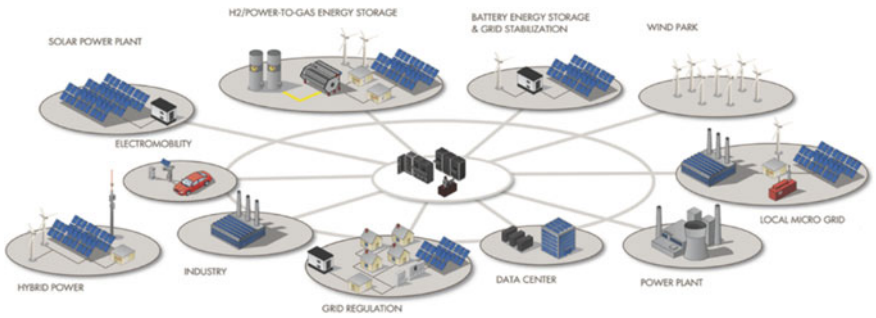


Fig. 1 Internet of energy

3.1 Relevance to Government Policies

This project is aligned with the Government's policy to focus on Research and Development in the Energy sector. It has become paramount for countries to be proactive in tackling the problems of global warming with efficient energy policies. Amongst the commitment made by the Malaysian Government to the 2015 Paris agreement was the voluntary reduction of up to 40% reduction of her Carbon emission intensity of GDP (Gross domestic product) with regards to the 2005 level [35]. Also, the country is aiming to reduce its green gas emissions for each unit of economic growth by 35% from 2005 levels by 2030; With international support, that could increase to 45%. Furthermore, part of the 11th Malaysian plan for encouraging sustainable energy use to support growth was improving the sustainability, reliability, and efficiency of the electricity sub-sector. The efficient management of electrical energy regulation of 2008 addresses the issue of energy efficiency by requiring more accountability from the big energy users of the industry. Some of the key points from the regulation require constant auditing of electrical energy management and proper in house policies to encourage efficient use of electrical energy such as the use of demand-side management [36]. Finally, the National Energy Efficiency Action Plan (2016–2020) encompasses all the other policies identified earlier. The action plan provides a detailed framework for how the country plans to reduce its energy consumption to meet international standards by 2025. And pivotal to this plan is the proper integration of demand-side management schemes to essentially help buildings monitor and effectively manage their energy consumption [37].

3.2 Proposed Deliverable

The IoT architecture has three levels; the control, cloud, and field level. The field level comprises of the old BAS and IoT layer. Various sensor modules are set in each room of a zone (three rooms constitute a zone) and they total estimations on

temperature, humidity, and CO_2. The measurements are transmitted through a wireless network, powered using ESP32. Created by Espress if Systems, ESP32 is a low-cost, low-power chip (SoC) device with Wi-Fi and dual-mode Bluetooth capabilities. The ESP32 was chosen because of its versatile communication modules (Bluetooth and Wi-Fi) and very low energy consumption.

These parameters are transmitted into the Zone Module in each region that in turn uses weather forecasts from Web servers or clouds together with measurements to predict the minimum cooling energy needed in each field. The Zone Module solves a problem of optimization using Machine learning. The measurement results are the cooling levels from which mass flow rates are predicted and transmitted via Wi-Fi to the IoT Gateway. Furthermore, an IoT layer provides the use of Ethernet-based communications at the application level, such as the Message Queue Telemetry Transport (MQTT) and Advanced Messaging and Queuing Protocol AMQP. They take the IP extensions of the IoT layer into account. Network extension services for interfacing IoT and BAS components with the network via IoT Gateway are supported by the IoT Layer (see Fig. 2 for the overall architecture of the proposed solution).

The sensor module has the purpose of sensing the temperature, CO_2, and relative humidity, and then transmitting measurements from several areas of a zone with wireless communication. In Fig. 2 you can see the sensor assembly. The main processor is an ESP32 with interfaces between the sensors. The ESP32 is chosen due to its low power consumption and its well-designed power management system is fairly stable. The temperature and humidity are measured using a DHT22 sensor. Similarly, the occupancy is detected using CO_2 sensor MH-Z19.

(1) Controller: The controller module consists of the latest Raspberry Pi Zero, an Arduino ESP32, a DHT22 (temperature and humidity sensor), and a CO_2 sensor MH-Z19. The Raspberry Pi Zero is used due to its ability to perform complex computations using Python or other programming languages and the Linux Operation System. More importantly, and ESP32 has wife capability that is used for logging data into the central repository through the IoT gateways with the cloud.

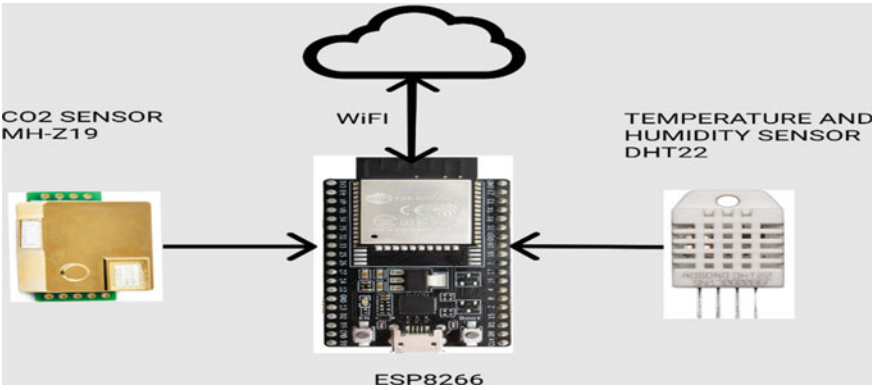


Fig. 2 Technical architecture

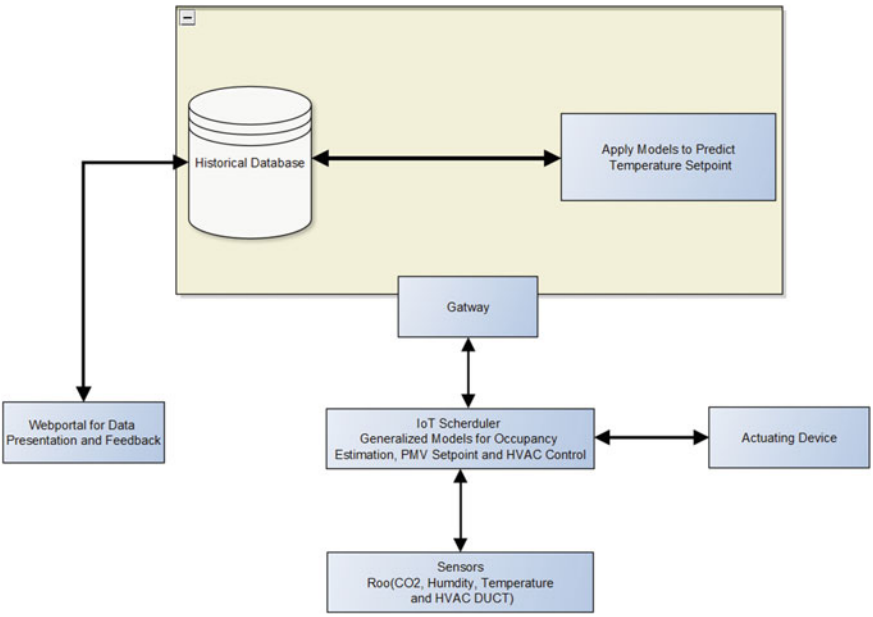


Fig. 3 The overall architecture of the proposed solution

(2) Software architecture: The IoT has numerous software components that function in different layers. The Raspberry pi at the field level is programmed using Python. The prediction of the occupant's comfort and the energy consumption is also done with Python with aid of open source libraries such as Scikit learn and Tensor flow. The IoT services are implemented on web-based applications (e.g., Node JS) for implementing MQTT clients. The BAS application that transfers data from BACnet and Lonworks is readily available from the vendors and no conversion is made.

The IoT gateway has interfaces to the different protocols using python and open-source software. The sensor node measurements implemented on an Arduino are transmitted using NRF24L01 to the RPi which aggregates it using python libraries (and AdafruitDHT) Then the RPi uses these measurements to compute the optimal cooling energy and transmits it to the central scheduler via Ethernet which is stored in a database using MySQL (See Fig. 3 for the overall architecture of the proposed solution).

4 Conclusion

This project aims to reduce the electrical energy consumption of HVACs in commercial buildings. Albeit recent government policies have made occupants and builders energy conscious, most of the existing buildings in Malaysia are not energy

efficient. For economic and cultural reasons, the best solution is to retrofit these buildings with energy monitoring and saving devices instead of demolition and reconstruction. We will leverage bi-directional communication provided by IoT to connect the network of HVAC in the building, and the subsequent big data is processed using powerful machine learning algorithms. The major reason for the huge energy consumption of centralized HVAC systems is the lack of proper monitoring. Primarily, the systems were not designed to provide feedback to building managers or operators. So a situation where a particular zone is over cooled will go unnoticed to the managers, leading to serious energy drain. Proper communication and actuation can eliminate this oversight. Energy consumption in buildings is dynamic, affected by different random variables. It is pure folly to simplify the energy consumption in buildings to binary problems or hardcoded into operational manuals. There is a direct correlation between building energy consumption and the random behavior of the occupants. Therefore, an effective BAS is smart and autonomous to reduce the energy consumption of a building while considering random occupant behaviors.

Acknowledgements This work was supported by the Universiti Tenaga Nasional (UNITEN) internal grant no RJO10517844/005 (2020). Many thanks to the Innovation & Research Management Center (iRMC), UNITEN who provided their assistance and expertise during the research.

References

1. Wang X, Feng W, Cai W, Ren H, Ding C, Zhou N (2019) Do residential building energy efficiency standards reduce energy consumption in China?—a data-driven method to validate the actual performance of building energy efficiency standards. Energy Policy 131:82–98
2. Pipattanasomporn M, Chitalia G, Songsiri J, Aswakul C, Pora W, Suwankawin S et al (2020) CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets. Sci Data 7:1–14
3. Gea, Global energy assessment—toward a sustainable future. Cambridge University Press, Cambridge, UK and New York, NY, USA and the International Institute for Applied Systems Analysis, Laxenburg, Austria, 2012
4. Hassan J, Zin R, Abd Majid M, Balubaid S, Hainin M (2014) Building energy consumption in Malaysia: an overview. Jurnal Teknologi 70
5. Wei T, Wang Y, Zhu Q (2017) Deep reinforcement learning for building HVAC control. In: Proceedings of the 54th annual design automation conference 2017, pp 1–6
6. Benedetti M, Cesarotti V, Holgado M, Introna V, Macchi M (2015) A proposal for energy services' classification including a product service systems perspective. Procedia CIRP 30:251–256
7. Mohandes SR, Zhang X, Mahdiyar A (2019) A comprehensive review on the application of artificial neural networks in building energy analysis. Neurocomputing 340:55–75
8. Curram SP, Mingers J (1994) Neural networks, decision tree induction and discriminant analysis: an empirical comparison. J Oper Res Soc 45:440–450
9. Hu M, Xiao F, Wang L (2017) Investigation of demand response potentials of residential air conditioners in smart grids using grey-box room thermal model. Appl Energy 207:324–335

10. Kang L, Liu L, An Q, Yang J, Zhao J, Jia H et al (2014) The equivalent thermal parameter model and simulation of air conditioner system in demand response programs. Energy Procedia 61:2004–2007
11. Hu M, Xiao F, Jørgensen JB, Wang S (2019) Frequency control of air conditioners in response to real-time dynamic electricity prices in smart grids. Appl Energy 242:92–106
12. Chen X, Wang J, Xie J, Xu S, Yu K, Gan L (2018) Demand response potential evaluation for residential air conditioning loads. IET Gener Transm Distrib 12:4260–4268
13. Wang J, Chen X, Xie J, Xu S, Yu K, Gan L (2019) Dynamic control strategy of residential air conditionings considering environmental and behavioral uncertainties. Appl Energy 250:1312–1320
14. Barrett E, Linder S (2015) Autonomous HVAC control, a reinforcement learning approach. In: Joint European conference on machine learning and knowledge discovery in databases, pp 3–19
15. Li B, Xia L (2015) A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings. IEEE Int Conf Autom Sci Eng (CASE) 2015:444–449
16. Jagarajan R, Abdullah MN, Asmoni M, Lee JY, Jaafar MN (2015) An overview of green retrofitting implementation in non residential existing buildings. Jurnal Teknologi 73
17. Husin S, Zaki N, Husain M (2019) Implementing sustainability in existing building through retrofitting measures. Int J Civil Eng Technol (IJCIET) 10:1450–1471
18. Borowski M, Mazur P, Kleszcz S, Zwolińska K (2020) Energy monitoring in a heating and cooling system in a building based on the example of the Turówka hotel. Energies 13:1968
19. Dixon T, McNamara P, Miller E, Buys L (2008) Retrofitting commercial office buildings for sustainability: tenants' perspectives. J Property Investment Finance
20. Oh TH, Pang SY, Chua SC (2010) Energy policy and alternative energy in Malaysia: issues and challenges for sustainable growth. Renew Sustain Energy Rev 14:1241–1252
21. Feng W, Wei Z, Sun G, Zhou Y, Zang H, Chen S (2020) A conditional value-at-risk-based dispatch approach for the energy management of smart buildings with HVAC systems. Electr Power Syst Res 188:106535
22. Ascione F, Borrelli M, De Masi RF, Vanoli GP (2020) Hourly operational assessment of HVAC systems in Mediterranean nearly zero-energy buildings: experimental evaluation of the potential of ground cooling of ventilation air. Renewable Energy
23. Jung HC, Kim JS, Heo H (2015) Prediction of building energy consumption using an improved real coded genetic algorithm based least squares support vector machine approach. Energy Build 90:76–84
24. Zhao HX, Magoulès F (2010) Parallel support vector machines applied to the prediction of multiple buildings energy consumption. J Algorithms Comput Technol 4:231–249
25. Papadimitriou T, Gogas P, Stathakis E (2014) Forecasting energy markets using support vector machines. Energy Economics 44:135–142
26. Kaytez F, Taplamacioglu MC, Cam E, Hardalac F (2015) Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines. Int J Electr Power Energy Syst 67:431–438
27. Al-Homoud MS (2001) Computer-aided building energy analysis techniques. Build Environ 36:421–433
28. Tso GK, Yau KK (2007) Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks. Energy 32:1761–1768
29. Ahmad A, Hassan M, Abdullah M, Rahman H, Hussin F, Abdullah H et al (2014) A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew Sustain Energy Rev 33:102–109
30. Zhao H-X, Magoulès F (2012) A review on the prediction of building energy consumption. Renew Sustain Energy Rev 16:3586–3592
31. Dounis AI, Caraiscos C (2009) Advanced control systems engineering for energy and comfort management in a building environment—a review. Renew Sustain Energy Rev 13:1246–1261

32. Fazenda P, Veeramachaneni K, Lima P, O'Reilly U-M (2014) Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems. J Ambient Intell Smart Environ 6:675–690
33. Less BD, Dutton SM, Walker IS, Sherman MH, Clark JD (2019) Energy savings with outdoor temperature-based smart ventilation control strategies in advanced California homes. Energy Buildings 194:317–327
34. Png E, Srinivasan S, Bekiroglu K, Chaoyang J, Su R, Poolla K (2019) An internet of things upgrade for smart and scalable heating, ventilation and air-conditioning control in commercial buildings. Appl Energy 239:408–424
35. Shahzad Y, Javed H, Farman H, Ahmad J, Jan B, Zubair M (2020) Internet of energy: opportunities, applications, architectures and challenges in smart industries. Comput Electr Eng 86:106739
36. Mahdi MN, Mohamed KS, Ahmad AR, Subhi MA (2020) The vision of 5G and cell-free communication networks in Malaysia. In: 2020 8th International conference on information technology and multimedia (ICIMU), pp 156–161
37. Mahdi MN, Ahmad AR, Ismail R, Subhi MA (2020) Improving big data technologies with visual faceted search. In: 2020 8th International conference on information technology and multimedia (ICIMU), pp 44–49

Features Extraction for Robust Face Recognition Using GLCM and CS-LBP

Abeer D. Salman, Mohammed Ahmed Talab, and Ruqayah R. Al-Dahhan

Abstract For fast and efficient automatic identity verification, biometric technology has developed rapidly. As one of the useful biometric identification technologies is face recognition. The applications of the face identification system-wide such as criminal detection, security verification, credit card verification, medicine, video conference, and other occasions. The main elements of the pattern recognition system include preprocessing, feature extraction, and classification. The feature extraction stage is concerned in this paper in order to suggest a robust algorithm useful in the classification stage. Many types of research try to recognize the face utilizing various algorithms, comparison among these algorithms is presented in this paper to show the effect of the feature extraction algorithm on human face classification. The main goal of this research is to combine Gray Level Co-Occurrence Matrix (GLCM) and Center-Symmetric Local Binary Pattern (CS-LBP) in a manner way to be used for extracting the texture feature of the human face classification. The experimental result explains that the proposed method given high accuracy in recognizing the human face where the proposed algorithm capable of recognizing 28 humans under different poses and illumination conditions.

Keywords Face recognition · Feature extraction · Texture features · GLCM · And CS-LPB

A. D. Salman (✉) · M. A. Talab
Computer Engineering Techniques, Al-Maarif University College, Ramadi, Iraq
e-mail: abeer.alnuaimi@uoa.edu.iq

R. R. Al-Dahhan
College of Computer Science and Information Technology, University of Anbar, Ramadi, Iraq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_16

1 Introduction

In recent years, for identity verification and to ensure that users provide information that is associated with a real person's identity, biometric identification becomes a critical requirement that needs to be fulfilled. Biometric identification technology has evolved rapidly. In particular, many patterns can be recognized such as criminals, credit cards, and security [1]. These patterns can be detected, verified, and monitored, respectively. Therefore, a pattern could be a word in cursive handwriting, a fingerprint image, a facial expression, or a voice signal. The goal of pattern recognition is to classify *objects* into numbers of *classes* [2, 3]. The classes can be either defined by supervised classification or unsupervised classification. In supervised learning, the objects are classified by the system designer while in unsupervised classification, based on the similarity of patterns, the classes are learned. Pre-processing, feature extraction, and classification are the three main processes in pattern recognition. Where a pattern consists of various features that differentiate individuals [4]. A feature vector is an n-dimensional vector that combines n features. The ability of differentiating various classes increases with the quality of a feature vector [5, 6]. One of the most essential and demanding technologies of biometric identification and pattern recognition is face recognition. In contrast with fingerprint, iris, and retina recognition, face recognition is a robust system that performs recognition under varying illumination and facial expression images and provides a strong layer of protection for securing facilities [7]. Based on the major processes of pattern recognition that are used for facial recognition, the dataset of images is gained and pre-processed to illuminate from noise. In addition, to improve the quality of images, histogram equalization is used, which in turn increases the contrast. This provides the best intensity gray-level distribution in the histogram. Feature extraction is the next phase, in particular, a set of feature vectors are extracted from the dataset. These vectors represent the original data and are used in the classification phase to define different classes based on the problem. Furthermore, before classification, a significant step in image processing is used which is image segmentation [8, 9].

The main goals of the presented research are improving and enhancing the statistical feature extraction method-based face recognition, where in this paper, feature extraction step is concerned to be used in the face recognition applications that are widely used nowadays. Some algorithms are used for this purpose, a comparative study on the global and local feature extraction method for face recognition are introduced in this research. To increase the efficacy of the recognition system, this research introduces a robust hybrid algorithm to extract the face features that are GLCM and Center-Symmetric Local Binary Pattern (CS-LBP).

This research is structured as follows: Sect. 2 describes the feature extraction definition and categories. The face recognition application is explained in Sect. 3. In Sect. 4, some previous works are presented. The proposed hybrid feature extraction algorithms and the implementation results are discussed in Sects. 5 and 6, respectively. Section 7 concludes the perspectives of the paper.

2 Feature Extraction

One of the major phases of pattern recognition and image processing is feature extraction, which is a remarkable dimension reduction form. In particular, the meaningful properties of the original, large input data are retained as a consequence of transformation into a restricted range of feature representations before being processed [10]. The learned models' accuracy is increased by extracting features of such data [8]. Measuring the values of an object to be distinct and recognized is the major goal of the feature extraction process. In computer vision, the challenge is how to extract optimal, distinguishable features of the images. Generally, texture, color, and shape are the most common features that are used to construct most image retrieval systems [11, 12]. To obtain high-quality features, three techniques have been used which are feature construction, feature selection, and feature extraction. In term of feature construction, input data are converted into a set of powerful features which are categorical, binary, or continuous. While the process of selecting informative features is feature selection. To extract beneficial features from images, Transform features, Spatial features, and Edge with boundary features are techniques that have been used for this purpose [10, 13].

2.1 Feature Types and Attributes

The most common types of features that can be extracted from images are color, shape, and texture features. These features are listed below:

(a) **Color features**

This visual type of object features results from transmission, emission, or reflection of light. The use of a specific color system or model that reproduces the representation of colors and turns them into numbers is called color space. Many color spaces have been utilized. For example, RGB refers to short Red, Green, and Blue, YCbCr Y is the brightness (luma) Cr is red, Cb is blue, and so on. A color feature is extensively used in image retrieval due to its intensity to the scaling, rotation, and translation of an image [8].

(b) **Shape features**

For the shape feature extraction, there are some characteristics used. These characteristics are perimeter, moment, orientation, and area. The external boundary of an object represents its outer shape extracting from other properties such as spatial properties, material composition, color, and so on. The process of extracting shape features is classified into two types, ignoring the interior of the shape and extracting shape features only from the boundary (Contour based) or using Region-based which calculates features from the entire region. Extracting shape features is

performed by using several algorithms such as Graham's Scan, convex hull, and Gift-wrapping [14].

(c) **Texture features**

A composition of sub-patterns provides complex visual patterns which is a texture feature. These features have particular characteristics such as slop, brightness, and size. Texture features need some statistics in multiple pixel areas. To extract the simplest texture features, some statistical properties need to be calculated from the image gray level histogram such as mean and variance. Another statistical calculation method used is gray-level co-occurrence matrices (GLCM) which are more powerful than the previous one. Such features have been quite popular in various texture analysis applications. However, they are sensitive to image illumination variations [15].

3 Face Recognition System

Machine recognition of human faces is a daunting task that is commonly used in video surveillance applications. However, it has significant challenges. These challenges are illumination, pose variation, images blurring, or face expression. Extracting effective features from the face region is a major step of robust face recognition. The main visual feature of an image is the texture that is considered the most powerful facial descriptors due to the robustness of extracting this feature to alterations of facial expression, pose, etc. [16–18]. Figure 1 shows feature face. In order to capture the general characteristics of textures and describe texture similarity, some texture descriptors are used. The first one is **Local Binary Pattern (LBP)**. This operator is efficient for texture classification. In particular, the image pixels are identified by the threshold for the neighborhoods of each pixel and the result is considered as a binary number. However, gaining satisfactory results cannot always be achieved by using the LBP descriptor, particularly in some applications. Therefore, another operator is utilized which is **Local Ternary Pattern (LTP)**. This descriptor is an extended version of LBP. Whereas LBP thresholds the pixels into 0 and 1, LTP thresholds pixels into three values using a threshold constant to figure out the difference between a pixel and its neighbours. However, LTP is limited and not effective in multiscale images. The third type of descriptors presented is **Median Binary Pattern (MBP)**. MBP thresholds the pixels of an image with respect to the median of the pixels within its neighbours. **Centre-Symmetric Local Binary Pattern (CS-LBP)** is another descriptor in which gray levels of the pixel's neighbours can be used to describe this pixel with ignoring the information of the center pixel. As a result, in some applications, this might affect the discriminative findings. Despite CS-LBP limitations, it is proven to be effective in face recognition under severe illumination variations and texture classification [19].

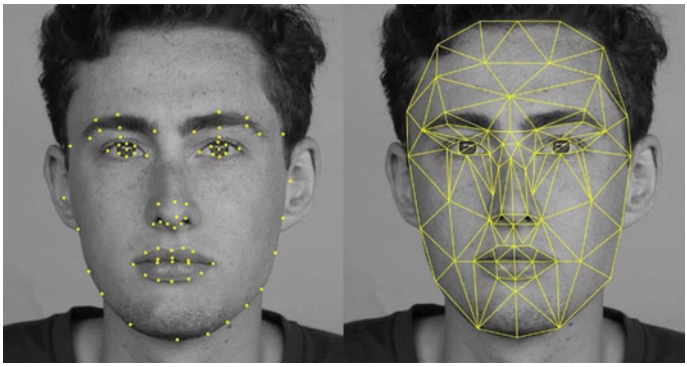


Fig. 1 Face features example

4 Literature Review

In this paper, some recent previous works that are recognized the face are surveyed in hence algorithm and data set used as well as the main results of their suggested works as explained in Table 1.

5 Methodology

This Section introduce new algorithm for creating face descriptor based on Center-Symmetric Local Binary Pattern (CS-LBP) and further aims to develop the new combine methods for face recognition. That exploration plans at take care of an issue of model, and surface acknowledgment dependent in measurable element. A goal of that exploration work is at utilize an improve strategy to separate highlights for the model and surface were a presentation of pictures during the framework as pictures and afterward removed primary concern double is then deciding the shape and surface passage highlights extricated from this figure. Proposed technique for highlight extract building up an improved element extraction strategy, all highlights of referenced dataset have been extricated using apply GLCM and CS-LBP strategy and perform characterization with one distinctive classifier called decision tree. The flowchart of this research is presented in Fig. 2. The input images are several face images to be recognized. Several stages constructed the proposed system that are: pre-processing stage for cropping to adjust the size of each image to be used.

The next stage is the combination of Center-Symmetric Local Binary Pattern (CS-LBP) and GLCM (Gray Level Co-occurrence Matrix) algorithms to extraction the texture features. In classification phase the decision tree classifier is used.

Table 1 Various works on face recognition

No	Author, year	Feature extraction algorithm	Results	Dataset used
1	Chowdhury et al. [17]	"Generalized two-dimensional Fisher's linear discriminate (G-2DFLD)"	High average recognition rates, it was varying between [92.98% and 98.58%]	Two databases are used: AT&T and UMIST1 The first one includes 400 images of 40 humans, where each person has ten images The second one consists of 575 images of 20 people
2	Rivera et al. [20]	Local principal texture pattern (LPTP)	The main result of this techniques is robust against rotation and illumination changes, high accuracy, and provides discriminant features The drawbacks of this algorithm are: give low accuracy in some cases, and LPTP effected to noise, slight rotation variations, and non-monotonic illumination	The databases used are Cohn-Kanade Facial Expression (CK) and Japanese Female Facial Expression (JAFFE)
3	Yu and Li [7]	Gray level co-occurrence matrix (GLCM)	The system could effectively recognize 2095 pairs of photos, where the recognition rate of the system is 99.8%	The name of the database don't mention, but the research randomly select 2099 pairs of face photos
4	Kapil [21]	"Discrete wavelet transform (DWT), discrete cosine transform (DCT), and local ternary pattern (LTP)" Equalized uniform local binary pattern (EULBP) has been implemented for the purpose of low-resolution images	The accuracy of DWT and DCT embedded with LTP is high, it reaches to i.e., 93.12% EULBP provide low performance	Not mention
5	Kim et al. [22]	LBP	The accuracy of the applied system is 95.11%	CASIA NIR-VIS 2.0 database (DB)

(continued)

Table 1 (continued)

No	Author, year	Feature extraction algorithm	Results	Dataset used
6	Zhang et al. [1]	LBP with Neural Network	The recognition system provides high accuracy and less sensitivity to noise	The dataset used consists of 500 positive samples and 2000 negative samples for training stage while the testing stage includes 200 positive samples and 400 negative samples
7	Wang et al. [23]	principal component analysis (2DPCA)	The applied 2DPCA based system no more difference of PCA due to the recognition rate and time is the same	ORL face database used in this system, it includes 400 samples for 40 people, 10 images for each one
8	Rajawat and Pandey [16]	DCT and GLCM	The accuracy of the proposed method reaches to 99.3%. Compared to the other methods, the method improved the performance by 0.5% for the non-face, and blur face	The used database contains 250 images grouped into three groups: faces, non-faces, and faces with noise
9	Kulkarni et al. [24]	LBP histogram	90 images are recognized accurately during the normal conditions but 80 images are recognized accurately under changing the illumination	In real time, the images are captured from webcam,100 images for each person
10	Sushama and Rajinikanth [25]	"Scale invariant feature transform (SIFT) and Dominant rotated local binary pattern (DRLBP)"	Just 75% is the accuracy of the introduced system	Not mention
11	Zhou et al. [19]	Improved center-symmetric local binary pattern (CS-LBP)	The proposed system better than the LBP, LTP, and CS-LBP	Four datasets are used, Olivetti research laboratory, Cambridge (ORL), UCSD computer vision, Yale university (YALE), extended Yale face dataset B (YALEB), and FERET

(continued)

Table 1 (continued)

No	Author, year	Feature extraction algorithm	Results	Dataset used
12	Kumar et al. [26]	The method called (GOL) is combination of GLCM and LBP	The proposed method improves the performance in terms of specificity, sensitivity, and retrieval time	Georgia Tech (GT-faces) and ORL-faces databases

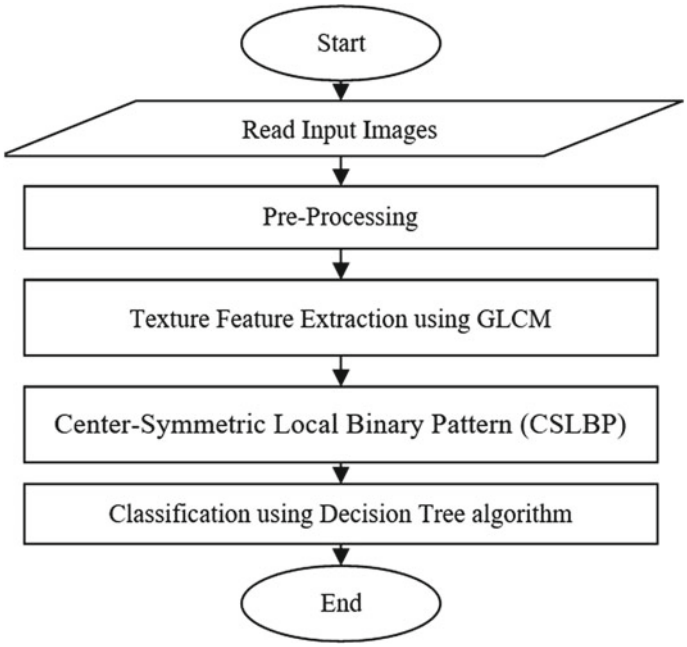


Fig. 2 Flowchart of the proposed recognition system

5.1 Gray Level Co-Occurrence Matrix (GLCM) Based Feature Extraction

The relationship between the two pixels that are separated by a certain space in the image making the occurrence of the gray scale distribution in the spatial position iterative that is formed the texture of image. The histogram of the gray image can be computed by enumerating a certain gray scale on a one pixel. If two pixels on the image that are hold a certain distance counted, GLCM is obtained. The GLCM or gray-level dependence matrix GLDM is a popular method describe the textures by explore the spatial correlation property of gray image [27]. The first step of GLCM

is creating co-occurrence matrices. CM collect spatial and statistical information, where texture metrics computed depending on this information. The distance d between two pixels and the orientation θ to the second pixel representing the spatial information. Each value in CM represented by $P(i, j, d, \theta)_{L \times L}$, where i and j are the row and column number in gray image respectively. L is the number of gray level, θ can be 0°, 45°,90°and 135° [28–30]. Figure 3 illustrates example of computing CM for 4 × 4 image with d = 1;

Because GLCM is only quantitative description of texture, it could not extract the texture image directly unless use the texture properties that are defined by Haralick, he introduced 14 statistical features that are [28, 31]:

1. *Mean* is the mean value of the gray level in an image.

$$\mu(I) = \sum_{i=1}^{G_{\max}} \{i.h_i\} \tag{1}$$

2. Variance σ reflects the changing domain of the gray level. Variance increase whenever the gray scale values oscillation from their means.

$$\sigma^2 = \sum_{i=1}^{G_{\max}} \left\{ (i-\mu)^2 \cdot h_i \right\} \tag{2}$$

3. *Contrast*: this feature represents the visibility of the textures, if the gray levels of each pixel pair are same, the contrast is low.

$$Contrast = \sum_{i=1}^{G_{\max}} \left\{ \sum_{j=1}^{G_{\max}} \left\{ (i-j)^2 \cdot p_{j,j} \right\} \right\} \tag{3}$$

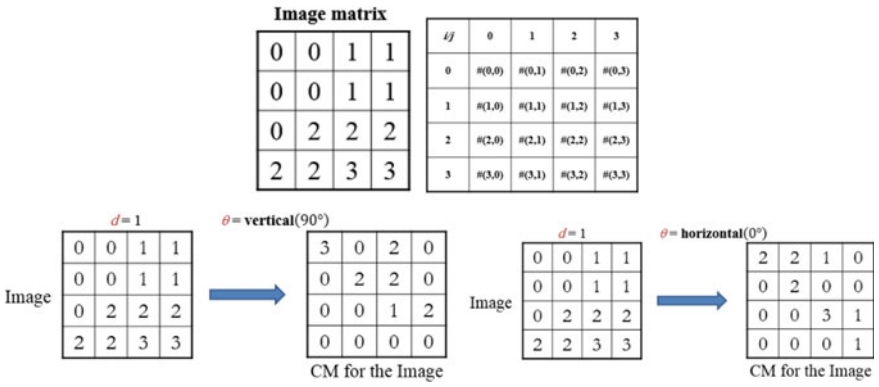


Fig. 3 GLCM example

4. *Homogeneity*: The similarity between the gray levels of each pixel pair leads to increase the homogeneity.

$$Homogeneity = \sum_{i,j=1}^{G_{\max}} \frac{p_{i,j}}{1+(i,j)^2} \tag{4}$$

5. *Correlation*: the linear dependence of the gray level to the neighborhood measured by correlation feature. If there are high dependence, the correlation is high.

$$Correlation = \frac{\sum_{i=1}^{G_{\max}} \left\{ \sum_{j=1}^{G_{\max}} \{(i-\mu_{I1})(j-\mu_{I2}).p_{i,j}\} \right\}}{\sigma_{I1}.\sigma_{I2}} \tag{5}$$

6. *Entropy*: this measure value close to zero if the image has no textures. Entropy measure the amount of information that is wanted for compressing the image [27].

$$Entropy = -\sum_{i=1}^{G_{\max}} \{h_i.\ln[h_i]\} \tag{6}$$

7. *The* similarity of gray image distribution and the thickness of the texture reflected by a feature named "Angular second moment (ASM)". ASM of the soft texture is lower than of coarse texture [27].

$$ASM = \sum_{i=1}^{G_{\max}} \left\{ \sum_{j=1}^{G_{\max}} \left\{ \left(p_{i,j}\right)^2 \right\} \right\} \tag{7}$$

8. *Energy*: calculate the number of iterated pairs. If there are high number of repeated pairs, the energy is high also.

$$Eng = \sum_{i=1}^{G_{\max}} \left\{ [h_i]^2 \right\} \tag{8}$$

9. *Inverse different moment (IDM)*: This feature measure the changes of the image texture, whenever IDM value was large, variation between image texture of different regions in the image was small and the locality is so regular [27].

$$IDM = \sum_{i=1}^{G_{\max}} \left\{ \sum_{j=1}^{G_{\max}} \left\{ \frac{1}{1+(i-j)^2} \cdot p_{i,j} \right\} \right\} \tag{9}$$

10. *Cluster Prominence (CP)*: this feature refers to degree of asymmetry. If the value of this feature is high, the image symmetric is less. While if the value is low, the peak value in GLCM matrix close to the mean values.

$$CP = \sum_{i=1}^{G_{\max}} \left\{ \sum_{j=1}^{G_{\max}} \left\{ (i + j - \mu_{I1} - \mu_{I2}) \cdot p_{i,j} \right\} \right\} \tag{10}$$

11. *Cluster Shade*: is the measurement of skewness of any matrix. The uniformity tested by this measure also. The result of this measure for asymmetric image is high.

$$s = \frac{1}{\sigma^3} \sum_{i=1}^{G_{\max}} \left\{ (i - \mu)^3 \cdot h_i \right\} \tag{11}$$

12. *Dissimilarity*: There are two features that are almost same except their weights, these features are contrast and dissimilarity. Same information is provided by these features but the range of values will be various. Measuring Contrast indicated by an exponential increase in the weights, while any linear increases in the weights refers to measure the dissimilarity.

$$Dissimilarity = \sum_{i=1}^{G_{\max}} \left\{ \sum_{j=1}^{G_{\max}} \left\{ |i - j| . p_{i,j} \right\} \right\} \tag{12}$$

where p be the normalized GLCM, n is the gray intensities number, h refers to the histogram.

5.2 Center-Symmetric Local Binary Pattern (CS-LBP) Based Feature Extraction

LBP is formed by the variance of local pixels. at a time, each pixel is regarded as a center pixel and each local surrounding pixel are comprised with that center pixel. A binary value is assigned to every surrounding pixel depending on the comparison. Specific weights multiply the binary values and summed up. The final summation value is named binary pattern value for that center pixel. this process repeated for all pixel image and the local binary map for the image obtained by all pattern value together. Histogram of this local binary map is computed for creating a feature vector. Mathematically, LBP defined as follows:

$$LBP_{P.R} = \sum_{s=0}^{P-1} S^2 \times T_1(I_s - I_c)$$
$$T_1(a) = \begin{cases} 1 & a \geq 0 \\ 0 & else \end{cases} \tag{13}$$

where P indicates the number of neighboring pixels while R refers to radius for neighboring pixels. Ic and Is are the Center pixel and surrounding pixels respectively. Final histogram of pattern map is computed by applying Eq. (2). LBP calculation is shown in Fig. 4 with a sample window.

$$H(L)|_{LBP} = \sum_{x_1=1}^{m} \sum_{x_2=1}^{n} T_2(LBP(X_1, X_2), L); \quad L \in \left[0, \left(2^P - 1\right)\right]$$
$$T_2(a_1, b_1) = \begin{cases} 1 & a_1 = b_1 \\ 0 & else \end{cases} \tag{14}$$

The modified version of LBP is CSPLB. For every pixel of the input image region, CSPLB extracts a local pattern. In CSLBP, each pixel is considered as center pixel like LBP. Difference of center symmetric pixels is calculated which is separate of the center pixel. Center symmetric local binary pattern value are generated by assigning four binary numbers to the center pixel based on that difference, the binary number are multiplied by four weights, and summed up to one value that is CSBPL value. Like LBP, the histogram is computed for feature vector of CSLBP map. The following equations explain CSLBP mathematically.

$$CSLBP_{P,R,T} = \sum_{s=0}^{\left(\frac{P}{2}\right)-1} 2^s \times T_3\left(I_s - I_{s+(P/2)}\right)$$
$$T_3(a) = \begin{cases} 1 & a > T \\ 0 & else \end{cases} \tag{15}$$

$$H(L)|_{CSLBP} = \sum_{x_1=1}^{m} \sum_{x_2=1}^{n} T_2(CSLBP(X_1, X_2), L); L \in [0, 15] \tag{16}$$

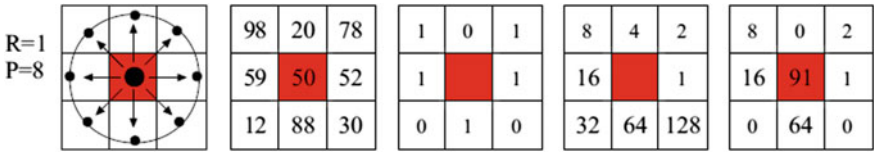


Fig. 4 LPB example

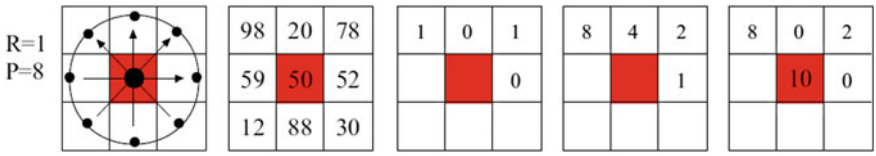


Fig. 5 CSLBP example

where T is the threshold value. An example of CSLBP is illustrated in Fig. 5. More details about CSLBP can be found in [32].

5.3 *Proposed Naïve Combination Vector Method*

The result of the combination of the feature extraction vectore[1] from GLCM with vectore[2] from CSLBP is one vector that is capable of calculating the new algorithm as shown in the equation:

$$f_{GLCM+CS-LBP} = f(f_{GLCM}, f_{CS-LBP}) \tag{17}$$

f value of vector.

6 Experiments and Results

The assessment of the proposed method is achieved by contrasting the proposed method with the base decision tree classifiers of the GLCM and LBP feature extraction methods.

This paper is the performance of the proposed method with the images from Yale Face Database that It comprises 16,128 photographs of 28 human subjects under 9 poses and 64 conditions of illumination. Figure 6 shows example of face images used.

Fig. 6 Examples Yale face database

Table 2 The classification results of proposed method proposed method, LBP and GLCM from 60 to 70% splitting of the training Yale Face Database

	LPB	GLCM	PROPOSED
60%	90.26	94.98	97.98
61%	91.25	92.03	98.52
62%	91.36	93.33	97.56
63%	92.65	91.53	98.56
64%	90.54	94.25	98.65
65%	92.54	90.76	98.65
66%	91.65	92.98	97.99
67%	90.48	93.69	98.56
68%	92.58	94.3	98.65
69%	93.5	94.65	97.56
70%	92.98	93.65	98.56
Average	**91.80**	**93.29**	**98.29**
Max	**93.5**	**94.98**	**98.65**

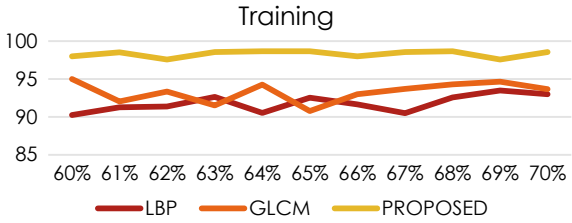


Fig. 7 Effects of the classification of the proposed system, LBP and GLCM using 6–70% of the training data (Yale Face Database)

The classification results of GLCM, LBP and proposed method from 60 to 70% splitting of the training Extended Yale Face Database B and the classification results of decision tree explained in Table 2. As shown from that table, the average recognition rate of the proposed system reaches to 98.29 compare with the other methods.

The classification results of proposed method, LBP and GLCM using 60% to 70% of the training data explained in Fig. 7.

The four experiments of the classification results of the 69% training sets with decision tree using GLCM, LBP and proposed feature extraction methods subsequently using the Extended Yale Face Database B dataset illustrated in Table 3 and Fig. 8.

Table 3 The four experiments of the classification results of the 68% training sets with decision tree using GLCM, LBP and proposed methods subsequently using the Yale dataset

Algorithm	1	2	3	4
LBP	93.5	93.69	92.65	92.85
GLCM	94.98	93.89	94.55	94.8
PROPOSED	98.65	98.79	99.89	98.54

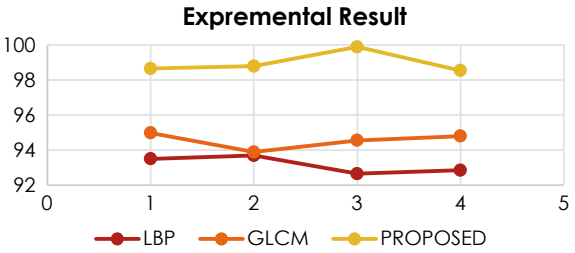


Fig. 8 The four evaluations of the classification outcomes of the 68% training sets with GLCM, LBP and proposed methods subsequently using the Yale B dataset

7 Conclusion

One of the most essential technologies of biometric identification and pattern recognition is face recognition. The goal of pattern recognition is to classify *objects* into numbers of *classes.* Pre-processing, classification, and feature extraction are the three main processes in pattern recognition. *In this research,* the most recently works that identifying the human face depending on various feature extracting algorithms surveyed. This paper introduced a new scheme for facial feature extraction by using combination of GLCM and CSLPB algorithms. The hybrid algorithm could be able to classify 28 humans depended on the features that are extracted by the suggested algorithm. The recognition rate reach to 98% using decision tree classifier.

References

1. Zhang H, Qu Z, Yuan L, Li G (2017) A face recognition method based on LBP feature for CNN. In: 2017 IEEE 2nd advanced information technology, electronic and automation control conference (IAEAC), pp 544–547. IEEE
2. Theodoridis S, Koutroumbas K (2009) Pattern recognition. Academic Press, 4th edn
3. Salman AD (2018) Review of the data mining algorithms with proposed hybrid theorem for increasing accuracy. J Al-Maarif University College 26:13–36
4. Sucharta V, Jyothi S (2013) A survey on various pattern recognition methods for the identification of a Different types of images. Research-Paper 4:178–181
5. Duda RO, Hart PE, Stork DG (2012) Pattern classification. John Wiley & Sons
6. Hussein AI, IH Abood, Salih WB (2014) Pattern recognition neural networks in classification groups of partial removable dentures lifetime. J Dent Oral Disorders Therapy

7. Yu J, Li C (2013) Face recognition based on euclidean distance and texture features. In: 2013 International conference on computational and information sciences, pp 211–213
8. Maksoud EAA, Barakat S, Elmogy M (2019) Medical images analysis based on multilabel classification. In: Machine learning in bio-signal analysis and diagnostic imaging. Elsevier, pp 209–245
9. Karyakarte S, Savant IJPR (2019) Pattern recognition process, methods and applications in artificial intelligence. 6(11)
10. Guyon I, Elisseeff A (2006) An introduction to feature extraction. In: Feature extraction. Springer, pp 1–25
11. Tian ping D (2013) A review on image feature extraction and representation techniques. 8(4): 385–396
12. Medjahed SA (2015) A comparative study of feature extraction methods in images classification. J Images Graph Signal Process 7(3):16
13. Talab MA, SNHS Abdullah, B Bataineh (2014) Edge direction matrixes-based local binar patterns descriptor for shape pattern recognition
14. Mingqiang Y, Kidiyo K, Joseph R (2008) A survey of shape feature extraction techniques 15(7):43–90
15. Sakai A et al. (2020) A method for the automated classification of benign and malignant masses on digital breast tomosynthesis images using machine learning and radiomic features 13(1):27–36
16. Rajawat A, Pandey MK (2017) Efficient feature extraction using hybrid face recognition method. In: 2017 7th international conference on communication systems and network technologies (CSNT), pp 158–162. IEEE
17. Chowdhury S, Sing JK, Basu DK, Nasipuri M (2010) Feature extraction by fusing local and global discriminant features: an application to face recognition. In: 2010 IEEE International conference on computational intelligence and computing research, pp 1–4. IEEE
18. Talab MA, Awang S, Ansari MD (2020) A novel statistical feature analysis-based global and local method for face recognition
19. Zhou N, Constantinides AG, Huang G, Zhang SJNC (2018) Face recognition based on an improved center symmetric local binary pattern. Neural Comput Appl 30(12):3791–3797
20. Rivera AR, Castillo JAR, Chae O (2012) Recognition of face expressions using local principal texture pattern. In: 2012 19th IEEE international conference on image processing, pp 2609–2612. IEEE
21. Kapil D (2015) Face recognition of blurred images using image enhancement and texture features. In: 2015 1st international conference on next generation computing technologies (NGCT), pp 894–897. IEEE
22. Kim H-I, Lee SH, Ro YM (2015) Multispectral texture features from visible and near-infrared synthetic face images for face recognition. In: 2015 IEEE international symposium on multimedia (ISM), pp 593–596. IEEE
23. Wang S, Wen G, Cai H (2017) Feature extraction and face recognition algorithm. In: 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 1910–1914. IEEE
24. Kulkarni O, Deokar S, Chaudhari A, Patankar S, Kulkarni J (2017) Real time face recognition using LBP features. In: 2017 International conference on computing, communication, control and automation (ICCUBEA), pp 1–5. IEEE
25. Sushama M, Rajinikanth E (2018) Face recognition using DRLBP and SIFT feature extraction. In: 2018 International conference on communication and signal processing (ICCSP), pp 994–999. IEEE
26. Kumar K et al (2020) Texture based FACE recognition using GLCM and LBP schemes. 13(13):1401–1411
27. Minghao L, Yifei L, Wenjuan H, Xiuxiu L (2019) Skin roughness evaluation method based on gray level co-occurrence matrix. In: 2019 Chinese control and decision conference (CCDC), pp 5671–5674. IEEE

28. Yu J (2010) Texture image segmentation based on gaussian mixture models and gray level co-occurrence matrix. In: 2010 Third international symposium on information science and engineering, pp 149–152. IEEE
29. Mahmood FH, WA Abbas (2016) Texture features analysis using gray level co-occurrence matrix for abnormality detection in chest CT Images 57(1A):279–288
30. Mohanaiah P, Sathyanarayana P, GuruKumar L (2013) Image texture feature extraction using GLCM approach. 3(5):1
31. Abbas Z, Rehman MU, Najam S, Rizvi SD (2019) An efficient gray-level co-occurrence matrix (GLCM) based approach towards classification of skin lesion. In: 2019 amity international conference on artificial intelligence (AICAI), pp 317–320. IEEE
32. Verma M, Raman B (2015) Center symmetric local binary co-occurrence pattern for texture, face and bio-medical image retrieval. J Visual Commun Image Represent 32:224–236

Toward a Full Exploitation of IoT in Smart Cities: A Review of IoT Anomaly Detection Techniques

Redhwan Al-Amri, Raja Kumar Murugesan, Eissa M. Alshari, and Hussam S. Alhadawi

Abstract The world is currently progressing towards a new connectivity era where billions of sensors are connected over a network called the Internet of Things (IoT). IoT enables a wide range of physical objects and devices to be connected and monitored with insufficient spatial and temporal detail. Despite their potential to improve multiple application domains, anomalies in the devices' behaviors pose a significant challenge, especially in the smart city's domain. Many research works have been devoted to determining such anomalous behaviors; however, there is a lack of comprehensive review focusing on anomaly detection techniques using statistical and machine learning methods in the smart cities domain. This work aims to fill this gap by presenting a review of anomaly detection techniques using statistical and machine learning methods. This paper explains the essential contexts related to IoT, followed by a review of the IoT anomaly detection techniques and their challenges, types, and detection modes. The paper then presents a summary of the related works related to smart cities. Finally, the open challenges and future directions were highlighted.

Keywords Internet of things (IoT) · Anomaly detection · Outliers · Statistical · Machine learning · Smart cities

R. Al-Amri (✉) · R. K. Murugesan
School of Computer Science and Engineering, Taylor's University, Kuala Lumpur, Malaysia
e-mail: redhwanmohammedabdullahalamri@sd.taylors.edu.my

R. K. Murugesan
e-mail: rajakumar.murugesan@taylors.edu.my

E. M. Alshari
Data Science Department, Universiti Malaysia Kelantan, Kelantan, Malaysia
e-mail: alsharieissa@umk.edu.my

H. S. Alhadawi
Computer Techniques Engineering Department, Dijlah University College, Baghdad, Iraq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_17

1 Introduction

The development and implementation of the internet of things (IoT) have provided different possibilities in technological improvement for various aspects of life. IoT is an essential component of the information communication technology (ICT) infrastructure of smart cities for ensuring environmental sustainability [1]. As per [2], the internet of things (IoT) is the future of the internet, where physical things and devices are connected via the internet. Another definition by Ullah et al. [3] stated that IoT is a mixture of sensors, actuators, and moving objects that offer real-time connectivity. Heterogeneous IoT sensors and devices are sources of a vast volume of data from various environmental factors, such as light, temperature, humidity, noise, power, voltage, electric current, etc. [4].

Numerous studies have investigated the techniques used for anomaly detection [5–9], and most of these techniques mainly addressed anomaly detection challenges associated with the use of machine learning approaches. However, these studies have not focused on other methods. This paper focuses on anomaly detection using statistical and machine learning techniques applied in the smart cities domain. There are three general anomaly detection modes: supervised, semi-supervised, and unsupervised. It is significant to mention that this study is different from the studies mentioned above in several practices; most of the existing studies have generally evaluated ML-based anomaly detection techniques while focusing less on the statistical methods. The main contributions of this paper are:

- Highlighting the different types of anomalies that exist in data.
- Reviewing the state-of-the-art techniques on anomaly detection in the smart cities domain.
- Highlighting the research challenges and prompting future directions.

This paper is organized as follows: Sect. 2 presents a comprehensive review of anomaly detection, including a detailed explanation of the challenges of anomaly detection in IoT, types, and detection modes. Section 3 offers a review of anomaly detection techniques in smart cities; this includes the nature of the data, machine learning techniques, the anomaly types, detection learning modes, domain applications, data sources, and finally, the evaluation criteria. Section 4 presents the challenges and future directions, while Sect. 5 is the study's conclusion.

2 Anomaly Detection

IoT devices nowadays produce large amounts of data streams that require immediate processing, especially in sensitive domains, such as medical care and environmental protection monitoring systems [10]. Each of the IoT devices sends continuous data based on the required application. With many active devices, typical data processing requires handling billions of incoming data daily [11]. Handling this volume of data

requires significant technical competence. Data on the computer itself is not actionable even when stored in a pre-processed form. The data gathered must be evaluated and analyzed to get actionable insights [12]. Anomaly detection is one of the analysis methods used to address such challenges and make IoT more robust and sustainable [13–15]. The main task is to identify the abnormal behavior of data that differs substantially from the expected or experienced pattern [12, 16].

Anomaly detection is a procedure mostly used to find odd or abnormal patterns of errors in wired and wireless networks that do not conform with the system's expected patterns [16]. In other words, it is a technique used to determine a weird pattern, called outliers, that does not conform with the predictable behaviour of the system [14]. This includes many business implementations, such as intrusion detection, smart city, healthcare system monitoring [17], credit card fraud detection, and malfunction detection in operating environments [18]. The anomaly detection concept has a broad meaning in the literature, pointing towards recognizing some observations or irregular actions that lead to doubts [4]. As stated by Mahdavinejad et al. [19], anomaly detection refers to the patterns or objects in a system that do not adapt with other expected patterns; these unusual patterns/behaviour are known as anomalies, outliers, novelties, exceptions, noise, surprises, or deviations. Solving an anomaly in IoT is challenging, yet, it is the primary step toward fully exploiting IoT in smart cities.

2.1 Types of Anomaly

In the data science literature, anomalies can be of three types depending on the environment; these are:

- Point anomaly

Where anomalies significantly differ from most other data points [20]. Figure 1a is an example of a point anomaly, where the anomaly differs from the rest of the data points.

- Contextual anomaly

Here, the context of the data points is anomalous and not the data point itself. As shown in Fig. 1b, the data point, t1, is identical to the data point, t2, but the latter point is considered anomalous because it appears in an anomalous context.

Fig. 1 Types of anomaly

- Collective anomaly

A collection of data points is considered anomalous not because of the data point themselves but because of the collection of these data points together. In Fig. 1c, the data points in the electrocardiogram are considered anomalies because of their appearance as a collection in this data and not because of the data points themselves.

2.2 Anomaly Detection Modes

Depending on the availability of training and testing data, as well as the characteristics of the traffic datasets used to create baselines, anomaly detection methods can be categorized into three modes, namely, supervised, semi-supervised, and unsupervised.

- **Supervised anomaly detection**

Supervised anomaly detection identifies anomalies based on the concept of creating a collection of labels that help in predicting future data [8].

- **Semi-supervised anomaly detection**

Semi-supervised anomaly detection is a technique that only models the normal data behavior while the other data patterns within the analysis are classified as anomalies [8].

- **Unsupervised anomaly detection**

Unsupervised anomaly detection focuses on unlabeled data and does not include a priori training and testing process. An example of this is the clustering of anomalies [8].

2.3 Anomaly Detection Techniques

This section addresses two main anomaly detection techniques, namely, statistical methods and machine learning methods:

- **Statistical methods**

Statistics refer to the process of data collection, structuring, definition, analysis, and interpretation [21]. The statistics domain could be viewed as a collection of resources that could be utilized to solve other problems. Statistical data analysis aims at predicting and extracting information. The algorithmic modeling culture approach is to identify the function that serves as interpretive variables for predicting the results. The data modeling culture considers that the variable response

could be measured as a function of predictor variables, parameters, and random noise [22].

- **Machine learning methods**

Machine learning (ML), as described in [23], is a study area that offers computers the capability to learn without being specifically programmed or trained. It is created to allow a system to learn from past or present experiences and use the information for future predictions [8]. While "learning" in ML is essential, it is not the goal; according to [24], the main purpose machine learning is to create a system that can identify significant patterns in the data automatically and accurately.

Machine learning algorithms could be generally divided into three groups based on properties, learning mode, and how data is used in particular; these are supervised, semi-supervised, and unsupervised algorithms [23]. This form of classification is critical in determining the function of the input data, the usefulness of algorithms and learning models with respect to their applications.

3 Anomaly Detection in Smart Cities

3.1 Anomaly Detection Using Statistical Methods

The smart city application in this study focuses on different sub-domains, such as healthcare systems, transportation, smart home, and smart objects. These application areas are vital in smart city management and the wellbeing of the citizens as they improve the quality of life and lead to the full exploration of IoT systems in smart cities. The following approaches and methods of IoT anomaly detection were discussed in this research using a statistical method.

1. **Intelligent Transportation System**

Intelligent transportation systems (ITS) cover a broad and diverse area in smart cities' domain application. Many techniques have been proposed for the detection of anomalies in ITS. The study by Han et al. [25] suggested a methodology based on ANOVA to detect a vehicle's anomalous behavior. The outcome of the research highlights that the existence of anomalies affects multiple data series; hence, correlation-based techniques were established to identify anomalous system behaviors.

A structured sparse subspace learning algorithm for the detection of abnormal behaviors was presented by He et al. [26]. To ensure the achievement of structured sparsity in the system, a structured norm was imposed on the projection coefficients matrix for efficient identification of anomaly sources. Experiments were conducted using two real-world flight datasets hosted at the UAV Laboratories of the University of Minnesota. During the experiments, the data used was only the flight data from takeoff to landing, and from the results, the proposed system performed better than some of the existing algorithms (Table 1).

Table 1 Anomaly detection using statistical methods in intelligent transportation systems

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
Structured sparse subspace learning [26]	2018	Continuous data	Contextual anomaly	Unsupervised learning on Thor Flight 107 and 111 flight data	Flight-critical sensors	Anomaly detection for flight safety	Two real-world flight datasets from UAV Laboratories	Accuracy and receiver operating characteristic (ROC)
Analysis of variance (ANOVA) [25]	2015	Continuous data	Collective anomaly	Supervised learning on the driving section	Vehicle data (engine, fuel, gear)	Abnormal behavior of the vehicle	Real commercial car dataset	Normal versus abnormal data

2. **Smart Objects**

Much attention has recently been given to the concept of smart objects as a reliable way of connecting numerous objects for easier communication. In smart cities, smart objects enable the collection of useful data and information from various domain applications. A latent correlation-based approach for detecting the abnormal behavior of a concrete pump truck was developed by Ding et al. [27]. The approach was developed for efficient anomaly detection based on data monitoring. The anomalies detection in this system was based on modelling the latent correlation between multiple correlative isomerous monitoring data series rather than relying on just one or more isomorphic monitoring data series. The modeling process was implemented using the latent correlation probabilistic model that can detect anomalies based on their relationships with the model. A total of 270 pump trucks was employed for the experiments in a supervised learning manner for anomalous behavior detection.

An unsupervised learning-based technique for the reduction of the annotation data burden into normal or abnormal patterns has been presented by Ho et al. [28]. The developed method ensured correct prediction by reducing the high variances of the mixture model. A smart nursing home with light switches was used for the experiments. The results showed the possibility of reducing the false alarm rate by at least 56% using the proposed method compared to the existing cluster-based frameworks that record low misdetection rates (Table 2).

3. **Health Care System**

Anomaly detection, analysis, and prediction are considered revolutionary steps that have redefined the healthcare systems, especially in the management of chronic diseases and remote monitoring for improved quality of life. In such scenarios, the major challenge is ensuring that the rate of false alarm generation is reduced. A dynamic time warping and density functions have been studied based on Photoplethysmogram (PPG) signals using a real dataset to monitor abnormal cardiac behavior [29]. The main novelty in this research is detecting and removing motion artefacts in PPG signals accurately and guarantee accurate detection of arrhythmias to minimize false-negative alarms. The results indicated that the proposed method performed effectively through a large set of experimentations with field-collected and MIT-Physionet PPG signals.

The study by Tonchev et al. [30] proposed a non-intrusive sleep analyzer for real-time sleep anomaly detection. The proposed method was based on combining non-invasive sensors and a two-stage sleep analysis algorithm, low and high-level reasoning using the data from sensors (i.e. Fitbit) and other health sensors. This also provides a chance to include devices from third parties that can monitor basic sleep behaviors using the analyzer, as well as to detect sleep outliers that can represent a vital pointer for mental and physical health.

Spectral coherence analysis was proposed by Pham et al. [31] for the identification of some anomalies in the sensory data of accelerometers. The practical approach here is spectral coherence analysis for determining anomalies scores for

Table 2 Anomaly detection using statistical methods in smart objects

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
Latent correlation probabilistic model (2015) [27]	2015	Continuous data	Collective anomaly	Supervised learning on patterns of concrete trucks	Engine speed and pump speed	Concrete pump trucks	Real dataset from 279 concrete pump trucks	Precision, recall, area under the ROC
Expectation maximization (2016) [28]	2016	Binary data	Point anomaly	Unsupervised learning on switches patterns	Light switches	Smart nursing homes	Own dataset	False alarm rate

FoG events. Using both the energy and freezing index functions, normal behavior is learned. Such a subject-independent method of identification of FoG is the main driver for a more efficient healthcare system for the elderly, particularly those with advanced Parkinson's illness. The developed device obtained an average mean sensitivity of 89.2 ± 0.3% (94 ± 0.3%) to ± SD (specificity) (Table 3).

3.2 Anomaly Detection Using Machine Learning Methods

Machine learning, as defined by Shanthamallu et al. [23], is a field of study that gives computers the ability to learn without being explicitly programmed. It is designed to allow a system to learn from the past or present and use the knowledge to make predictions or decisions for the future [8]. Although "learning" is significant in machine learning, it is not the goal; the primary purpose of machine learning, as per [24], is to produce a system that can automatically and correctly detect meaningful patterns in data. Numerous IoT anomaly detection studies have been proposed in smart cities, yet, researchers are still focusing much on this area in a bid to devise better solutions.

1. **Intelligent Transportation System**

In intelligent transport, machine learning models are helpful in the detection and prediction of abnormal behaviors. For instance, the study by Janakiraman and Nielsen [33] recently introduced the Extreme Learning Mechanism (ELM) for addressing the issue of high computational training time over large datasets. A useful generalization was also provided to develop a scalable anomaly detection technique over large datasets. Experiments were conducted using a real aviation safety benchmark problem that contains 43,000 flight data (comprising of radar information, the distance of nearby aircraft, and flight trajectories). Furthermore, deep learning models have also found application in different fields, such as in speech classification, image recognition, and natural language processing, owing to their ability to detect anomalies; hence, many in-depth learning-based solutions have been proposed and developed by many researchers.

A reinforcement learning method for detecting anomalies in the motor of an unmanned aerial vehicle based on temperature data has been developed by Lu et al. [34]. Most of the reinforcement learning algorithms typically rely on the reward mechanism; they ensure efficient control of the motor temperature by causing increases in the speed when the range of temperature is appropriate. As the temperature increases, the rate is reduced to avoid damages to the motor, which is the primary cause of drone crashes.

The study by Farshchi et al. [32] presented a regression-based model for finding the contextual anomalies in air traffic control systems via the determination of the events-resource metrics correlations in the logs files. They also provided further details of possible improvements by suggesting the inclusion of change detection frameworks and time windows utilization on contextual anomalies. Experiments

Table 3 Anomaly detection using statistical methods in healthcare system

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
Dynamic time warping (DTW) and density function [29]	2016	Continuous data	Contextual anomaly	Supervised learning on Physionet PPG signals	(PPG) signals	Cardiac behavior	N/A	Precision, recall specificity
Hidden Markov model [30]	2016	Continuous data	Contextual anomaly	Unsupervised learning on eWALL Consortium parameters	Wearables Fitbit	Sleep analyzer	Own dataset	No performance measure
Regression model [32]	2018	Continuous data	Contextual anomaly	Supervised learning on historical data	Software logs	Air traffic control systems	Own dataset	Sensitivity

were conducted over three failure settings in which the first setting captures a working system, but the services are unresponsive. The second setting is an operating system, but the service is unresponsive due to the prolonged waiting time for an event. The third setting is a system crashed system services were unexpectedly terminated.

In the study by Bose et al. [35], the successful application of a support vector machine (SVM) algorithm was reported for the provision of real-time instructions and warnings to drivers regarding the condition of roads to ensure a safe driving experience. In the proposed system, events were categorized based on the detection of driving activities, such as braking and acceleration, as well as road anomalies, such as potholes and bumps; this categorization was done using SVM while a locally running Fast Dynamic Time Warping (FastDTW) algorithm was used for the provision of real-time information to the drivers.

The development and image-based training of a Deep Convolutional Neural Network (CNN) for the detection of anomalies on railway track has been presented by Gibert et al. [36]. CNN-based approaches are mainly identified by their network structure, weight update mechanism, and activation function. The experiments in this study were conducted using images captured via 85 miles of continuous tracking in a supervised training mechanism. The system employed multiple detectors in a multitask learning framework for better anomaly detection accuracy on railway ties & fasteners.

An unsupervised deep learning-based autoencoder has been presented by Luo et al. [37]; the study also demonstrated a stacked denoising autoencoder model for unsupervised learning of the robust feature space. Then, the output of the autoencoder was used as the input to the Gaussian distribution-based anomaly detection algorithm. Experiments were conducted on the turbofan gas turbine engine of a civil plane for the detection of anomalous engine behavior. The proposed method achieved better anomaly detection accuracy compared to the conventional methods; hence, the proposed method was recommended for application in engine health management.

A similarity metric-based CNN was developed by Lang et al. [38] to detect traffic anomalies. The method comprised of (i) a Siamese network with a hierarchical attention model via word embedding for effectively measuring similarities between anomalies and the templates; (ii) a deep transfer learning method for auto annotation of an unlabeled set while ensuring fine-tuning of the network; (iii) a background modelling technique that combines temporal and spatial information for anomaly extraction. The experimental results showed that the proposed approach performed better in terms of accuracy than the deep convolutional generative adversarial network (DCGAN) and the AutoEncoder by 3% and 5%, respectively.

An anomaly detection approach for smart city application has been proposed by Santos et al. [15] using Birch clustering and RC outlier; the proposed system focused on low-power Fog Computing solutions and was tested for performance on the Antwerp's City of Things testbed. The outcome of the experiments showed that

both mechanisms of anomaly detection (Birch clustering and RC outlier) are implementable on fog resources that are close to IoT sensors and can send timely alerts during emergencies (Table 4).

2. **Smart Objects**

Clustering and rule-based approaches are machine learning techniques developed to predict and detect an anomaly in smart object systems. Hybrid methods have been developed by Islam et al. [39] via hybridization with fuzzy logic regarding the rule-based approaches. The study relied on the calculated level of belief to address ambiguities, vagueness, and imprecision. The prediction of anomalous behavior was based on water level monitoring using a rain gauge and temperature sensors.

The study by Sun et al. [40] employed a novel way of considering rules for domain knowledge-based verification of the rules' exactness. For the experiments, a school building was considered with several sensors to detect abnormal levels of energy consumption to ensure a comfortable living with reduced energy utilization. Rule-based models are mainly beneficial due to the human understanding of their output. This made them more applicable in areas where there is a need for ML outputs that a human can understand.

Researchers in [41] presented a new cluster heat map approach for the monitors for office environments' energy usage. The experiments were performed with 127 sensors deployed in two-floor buildings to measure the total power, phase angle, reactive power, fair value, voltage, current, motion, and temperature readings. The experiments' results showed that the proposed algorithm could identify anomalies and capture the cluster structures' changes within the streaming data.

Furthermore, the study by Chen et al. [42] developed temporal and spatial-temporal methods for anomaly detection in environmental datasets. The study also presented a visualization method for anomaly prediction in monitored systems. The evaluation of the system showed that it could identify data outliers and detect anomalous patterns recognized by the government authorities and the general public.

The use of the ANN algorithm for the development of an auto anomaly detection technique in heterogeneous sensor networks has been presented by Cauteruccio et al. [4]; the approach was based on cloud data and coupling edge data analyses. Experiments were conducted on real data sourced from an indoor building that was distorted with a range of artificial impairments. The experimental results showed the self-adaptation capability of the proposed method to environmental changes, as well as its effectiveness in anomaly detection.

The study by Antonini et al. [43] reported designing a system for smart audio sensors using the elliptic envelope and isolation forest; the developed method can record and pre-process raw audio data streams into audio data features before wirelessly propagating them to a modular IoT gateway. A real smart environment was deployed for the training of the affordable IoT gateway in the detection of abnormal sound events within an office environment (Table 5).

Table 4 Anomaly detection using machine learning methods in intelligent transportation systems

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
Extreme learning machines [33]	2016	Continuous data	Collective anomaly	Supervised learning on 43,000 flights data	Radar, aircraft trajectories and nearby aircraft distance	Aviation safety	Real aviation safety benchmark dataset	Area under the curve
Reinforcement learning [34]	2017	Continuous data	Point anomaly	Simi-supervised learning on drone captured data	Temperature sensor	Unmanned aerial vehicles	Own simulations dataset	Accuracy
Regression model [32]	2018	Continuous data	Contextual anomaly	Supervised learning on historical data	Software logs	Air traffic control systems	N/A	Precision, recall and accuracy
Super vector machine [35]	2018	Continuous data	Contextual anomaly	Supervised learning on historical data	GPS and accelerometer	Driving patterns	Own dataset	Accuracy
Deep convolutional neural network [36]	2018	Image data	Point anomaly	Supervised learning on CTIV platform	Camera	Railway track	Own real dataset	Accuracy
Autoencoders [37]	2017	Continuous data	Point anomaly	Unsupervised learning on Civil plane data	Gas turbine engine	Flights	Own simulation dataset	Precision, recall, F1 score
Convolutional neural network (CNN) [38]	2019	Continuous data	Point anomaly	Supervised learning on abnormal events	Image sensors	Traffic control	DCGAN dataset	Accuracy = 0.953
Birch clustering and RC outlier [15]	2018	Continuous data	Point anomaly	Unsupervised learning on air quality monitoring	Multi-radio sensor and air quality sensors	Air quality monitoring	Own real dataset	N/A

Table 5 Anomaly detection using machine learning methods in smart objects

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
Rule-based system [39]	2018	Continuous and image data	Collective anomaly	Supervised learning	Temperature, humidity, light intensity, audio and image	Efficient energy consumption in school building	Breast cancer Wisconsin dataset	Number of conflicts and miss detection rate
Belief-rule-based association rules [40]	2016	Continuous and image data	Collective anomaly	Supervised learning on patterns of incomplete vague data	Temperature and rain gauge	Flood prediction	N/A	Area under the curve
Clustering [41]	2016	Continuous and image data	Contextual anomaly	Unsupervised learning on patterns of smart cities IoT data	Power, temperature, motion sensor, and vibration	Smart city	Own real dataset	Normal versus abnormal comparisons
Spatial-temporal framework [42]	2017	Continuous and image data	Collective anomaly	Simi-supervised learning on patterns of smart cup usage	liquid level sensor, and force sensing resistors	Smart cup	N/A	No comparison
Artificial neural network [4]	2019	Continuous and image data	Point anomaly	Unsupervised learning on patterns of WSN nodes	WSN nodes	Smart building	N/A	Accuracy

3. **Healthcare System**

The use of the HEAL model to develop a data-driven approach over binary data for contextual anomalies detection over the tap in home environments has been reported by Manashty et al. [44]. The aim of developing the system was to facilitate quick and accurate health anomalies prediction by bridging the gap between symptoms and diagnosis trend data.

The study by Lyu et al. [45] reported the development of a Fog-Empowered anomaly detection model using an efficient hyper ellipsoidal clustering algorithm; this study exploited the processing capability of the Fog computing platform. The clustering and anomaly detection tasks were performed by the nodes of the Fog and Cloud layers, thereby facilitating quicker anomaly detection. The performance of the system was evaluated on both real and synthetic datasets and proven to significantly reduce both energy consumption and latency when compared to the rest of the Distributed and Centralized schemes. However, the detection accuracy of the proposed system was comparable to that of a Centralized scheme.

The detection of performance anomalies on real-time infrastructure metrics that were collected via continuous monitoring of workflow tasks for resource consumption using Hierarchical Temporal Memory (HTM) has been reported by Rodriguez et al. [46]. The proposed system can process measurement streams in both online and unsupervised manners and can adapt to changes in the underlying data statistics. The proposed model was shown to be capable of accurately identifying the performance anomalies on varying resource consumption metrics as a result of numerous competing workloads that have been added to the system.

The development of an IoT-based long-term monitoring system using Bayesian networks-based methods has been reported by Azimi et al. [47] for objective maternal sleep quality assessment. Focusing on point anomaly detection using a semi-supervised machine learning approach, the authors evaluated the trained model using a real dataset collected from 13 pregnant women (Table 6).

4. **Smart Homes**

In the smart homes domain, Bayesian techniques have been used by Gagewski et al. [17] to develop a two-tier IDS that relies on ML methods for the classification of the monitoring records, as well as for searching potential abnormal activities with a network at the data center of the service provider. From the results, the combination of local anomaly detection with anomaly correlation by the service providers can offer reliable info on the most frequent misbehaving IoT devices because of an infection. The study by Deng et al. [14] reported the detection of outliers in large-scale IoT sensor data using the One-Class Support Tucker Machine and Genetic Algorithm. The performance of the proposed approach was evaluated on real datasets. From the results, there were significant improvements in the efficiency and accuracy of anomaly detection, as well as retention of the intrinsic structure of big sensor data.

Anomaly detection in a large-scale Smart Home dataset using Neural networks was reported by Legrand et al. [49]. In this study, two types of auto-coders that are

Table 6 Anomaly detection using machine learning methods in healthcare system

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
HEAL model [44]	2015	Binary data	Contextual anomaly	Supervised learning on smart home system	Tap sensors on the objects	Elderly homes	Tim Van Kastern's public datasets [48]	No performance measures
Hyper ellipsoidal clustering algorithm [45]	2017	Continuous data	Point anomaly	Unsupervised learning on energy consumption	Waspmote sensor	Healthcare	Two synthetic and two real-world datasets	N/A
Hierarchical temporal memory (HTM) [46]	2018	Continuous data	Point anomaly	Unsupervised learning	N/A	Bioinformatics field	Two real scientific workflows	N/A
Bayesian networks-based [47]	2019	Continuous data	Point anomaly	Simi-supervised learning on maternal sleep	Wearable sensors	Pregnancy monitoring	Real dataset of 13 pregnant women	Simple aggregate method

sensitive to various forms of anomalies were selected. Different convolutional and recurrent models were also trained on a set of high-dimensional academic data. The authors evaluated the results by proposing two indices that exploit the approximation and generalization capabilities of neural networks. A review of the obtained results allowed the determination of the recurrent auto-encoders with a better generalization capability and the most effective approach for anomalies detection in the utilized data set.

Researchers in [12] proposed using Density-based Spatial Clustering of Applications with Noise (DBSCAN) clustering algorithm for the development of an anomaly detection system; the system was proposed for efficient management of security in IoT environments. The proposed approach was evaluated and observed to achieve higher detection accuracy and lower computational complexity (Table 7).

4 Open Challenges

The smart city paradigm continues to evolve with innovation and deployment that are limited only to the developed nations. Singapore as an example has taken measures to use data from citywide IoT sensors to track daily living. Singapore's Smart Nation Program (SNP) has integrated current technologies and existing infrastructure for connecting all communities online. Nevertheless, anomalies are one of the major challenges faced within the IoT environment. Therefore, further studies on anomaly detection within smart cities are highly recommended. Generally, many existing challenges need to be solved within the domain of IoT anomaly detection in smart cities. It was noticed that the real datasets of smart cities are difficult to access and require a complex process to access the running system data. There is a huge gap in formalizing a means of accessing data logs and sensory data flow, creating a model, and validating it in smart cities' real-life environments.

Several statistical approaches have been well established in recent decades to detect anomalies. Considering the noticeable increase in the installation of tap sensors on IoT Smart items, such validated methods currently face a major challenge. Monitoring devices are now fitted with sophisticated hardware and software and the number of parameters has increased significantly to include the data source. Thus, statistical approaches are no longer capable of handling such data volumes.

On the other hand, machine learning techniques were developed for dealing with high-dimensional data and detecting abnormal patterns within the system. In the other fields, such as computer vision and speech processing, a subsection of ML is a discipline named deep-learning which has been successful in generating more accurate results from complex datasets. There is a research gap in applying new models and evaluating their anomaly detection capabilities in different fields, especially those used in smart cities, healthcare systems, intelligent transport, manufacturing, and smart objects.

Table 7 Anomaly detection using machine learning methods in smart homes

Methods	Year	Nature of the data	Anomaly type	Anomaly detection modes	Type of sensors	Application area	Dataset	Evaluation criteria
Bayesian [17]	2019	Continuous data	Point anomaly	Supervised learning on smart metering	WSN sensors	Smart home	Own simulation dataset	N/A
One class support tucker machine and genetic algorithm [14]	2019	Continuous data	N/A	Unsupervised learning on large-scale IoT sensor data	N/A	Smart building	Own real dataset	N/A
Neural networks [49]	2019	Continuous data	Contextual anomaly	Supervised learning on home sensors	20 Home sensors	Smart home	(REFIT) Smart Home dataset [6]	N/A
Density-based spatial clustering of applications with noise (DBSCAN) [12]	2019	Continuous data	Point anomaly	Unsupervised learning on LAB network	N/A	Lab network	Own simulation dataset	Accuracy = 91.5%

Furthermore, data complexity, such as imbalanced datasets, unexpected noises, and redundancy within the data, is one of the main challenges faced when it comes to developing an anomaly detection model for use in smart cities. Well-developed approaches are necessary for curating the datasets and collection of useful information and knowledge. While current technology can detect outliers, the results are still unreliable due to accuracy-related problems. In certain circumstances, smart cities achieve higher accuracy at the expense of high computational power and time, which is one of the main challenges in the smart cities domain as the devices are resource-constrained.

5 Conclusion

Anomaly detection for an important application domain such as smart cities is an incredibly critical issue. A main finding in anomaly detection is that it is not an issue that is well formulated. Anomaly detection will decrease the operating threats and prevent unseen complications and downtime of the systems. This paper reviewed IoT anomaly detection techniques in smart cities, focusing on the trends and open challenges using two of the main anomaly detection techniques: statistical and machine learning techniques. Twenty-eight recent research papers were reviewed and evaluated within the smart cities' domain. The investigation of different use cases in this domain helped identify the challenges associated with IoT anomaly detection in smart cities. Furthermore, the state-of-the-art approaches and the related challenges in IoT anomaly detection in smart cities were highlighted, followed by a discussion on the open challenges for future research. As future work, the focus is drawn toward investigating the anomaly detection techniques on machine learning over the data stream, characterized as being infinite and evolving.

References

1. Bibri SE (2018) The IoT for smart, sustainable cities of the future: an analytical framework for sensor-based big data applications for environmental sustainability. Sustain Cities Soc 38:230–253. https://doi.org/10.1016/j.scs.2017.12.034
2. Arowoiya VA, Oke AE, Aigbavboa CO, Aliu J (2020) An appraisal of the adoption internet of things (IoT) elements for sustainable construction. J Eng Des Technol. https://doi.org/10.1108/JEDT-10-2019-0270
3. Ullah F, Al-Turjman F, Nayyar A (2020) IoT-based green city architecture using secured and sustainable android services. Environ Technol Innov 101091. https://doi.org/10.1016/j.eti.2020.101091
4. Cauteruccio F et al (2019) Short-long term anomaly detection in wireless sensor networks based on machine learning and multi-parameterized edit distance. Inf Fusion 52(November 2018)13–30. https://doi.org/10.1016/j.inffus.2018.11.010
5. Salehi M, Rashidi L (2018) A survey on anomaly detection in evolving data [with application to forest fire risk prediction]. SIGKDD Explor Newsl 20(1):13–23

6. Donevski M, Zia T (2019) A survey of anomaly and automation from a cybersecurity perspective. In: 2018 IEEE Globecom Work. GC workshops 2018—proceedings, pp 1–6. https://doi.org/10.1109/GLOCOMW.2018.8644456
7. Behniafar M, Nowroozi A, Shahriari HR (2018) A survey of anomaly detection approaches in internet of things 10:2
8. Ariyaluran Habeeb RA, Nasaruddin F, Gani A, Targio Hashem IA, Ahmed E, Imran M (2019) Real-time big data processing for anomaly detection: a survey. Int J Inf Manage 45 (September 2018):289–307. https://doi.org/10.1016/j.ijinfomgt.2018.08.006
9. Fahim M, Sillitti A (2019) Anomaly detection, analysis and prediction techniques in IoT environment: a systematic literature review. IEEE Access 7:1–1. https://doi.org/10.1109/access.2019.2921912
10. Lee I (2017) Big data: dimensions, evolution, impacts, and challenges. Bus Horiz 60(3):293–303. https://doi.org/10.1016/j.bushor.2017.01.004
11. Alaba FA, Othman M, Hashem IAT, Alotaibi F (2017) Internet of things security: a survey. J Netw Comput Appl 88:10–28
12. Qin T, Wang B, Chen R, Qin Z, Wang L (2019) IMLADS: intelligent maintenance and lightweight anomaly detection system for internet of things. Sensors (Basel) 19(4). https://doi.org/10.3390/s19040958
13. Wu M, Song Z, Moon YB (2019) Detecting cyber-physical attacks in cyber manufacturing systems with machine learning methods. J Intell Manuf 30(3):1111–1123. https://doi.org/10.1007/s10845-017-1315-5
14. Deng X, Jiang P, Peng X, Mi C (2019) An intelligent outlier detection method with one class support tucker machine and genetic algorithm toward big sensor data in internet of things. IEEE Trans Ind Electron 66(6):4672–4683. https://doi.org/10.1109/TIE.2018.2860568
15. Santos J, Leroux P, Wauters T, Volckaert B, De Turck F (2018) Anomaly detection for smart city applications over 5G low power wide area networks. In: IEEE/IFIP Networking operation and management symposium: cognitive management a cyber world, NOMS 2018, pp 1–9. https://doi.org/10.1109/NOMS.2018.8406257
16. Vilenski E, Bak P, Rosenblatt JD (2018) Multivariate anomaly detection for ensuring data quality of dendrometer sensor networks. Comput Electron Agric 162(November 2018):412–421. https://doi.org/10.1016/j.compag.2019.04.018
17. Gagewski M, Batalla JM, Levi A, Togay C, Mavromoustakis CX, Mastorakis G (2019) Two-tier anomaly detection based on traffic profiling of the Home automation system. Comput Netw 158:46–60. https://doi.org/10.1016/j.comnet.2019.04.013
18. Ratasich D, Khalid F, Geissler F, Grosu R, Shafique M, Bartocci E (2019) A roadmap toward the resilient internet of things for cyber-physical systems. IEEE Access 7:13260–13283. https://doi.org/10.1109/ACCESS.2019.2891969
19. Mahdavinejad MS, Rezvan M, Barekatain M, Adibi P, Barnaghi P, Sheth AP (2018) Machine learning for internet of things data analysis: a survey. Digit Commun Netw 4(3):161–175. https://doi.org/10.1016/j.dcan.2017.10.002
20. Chandola V, Banerjee A, Kumar V (2009) Survey of anomaly detection. ACM Comput Surv 41(3):1–72. https://doi.org/10.1145/1541880.1541882
21. Snedecor GW, Cochran WG (1967) Statistical methods, ed 6, Ames, Iowa, 1967. Iowa State Univ Press Sect 12:349–352
22. Breiman L et al. (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat Sci 16(3):199–231
23. Shanthamallu US, Spanias A, Tepedelenlioglu C, Stanley M (2017) A brief survey of machine learning methods and their sensor and IoT applications
24. Shalev-Shwartz S, Ben-David S (2014) Understanding machine learning: from theory to algorithms. Cambridge university press
25. Han ML, Lee J, Kang AR, Kang S, Park JK (2015) A statistical-based anomaly detection method for connected cars in internet, 89–97. https://doi.org/10.1007/978-3-319-27293-1

26. He Y, Peng Y, Wang S, Liu D, Leong PHW (2018) A structured sparse subspace learning algorithm for anomaly detection in UAV flight data. IEEE Trans Instrum Meas 67(1):90–100. https://doi.org/10.1109/TIM.2017.2754698
27. Ding J, Liu Y, Zhang L, Wang J, Liu Y (2015) An anomaly detection approach for multiple monitoring data series based on latent correlation probabilistic model. Appl Intell. https://doi.org/10.1007/s10489-015-0713-7
28. Ho C, Chou C, Chien Y, Lee C (2016) Unsupervised anomaly detection using light switches for smart nursing homes. https://doi.org/10.1109/DASC-PICom-DataCom-CyberSciTec.2016.139
29. Puri C (2016) iCarMa: inexpensive cardiac arrhythmia management—an IoT healthcare analytics solution 3–8
30. Tonchev K, Koleva P, Manolova A, Tsenov G, Poulkov V (2016) Non-intrusive sleep analyzer for real time detection of sleep anomalies 400–404
31. Pham TT, Nguyen DN, Dutkiewicz E, Mcewan AL, Leong PHW (2017) Wearable healthcare systems: a single channel accelerometer based anomaly detector for studies of gait freezing in parkinson's disease
32. Farshchi M, Weber I, Della Corte R, Pecchia A, Cinque M (2018) Contextual anomaly detection for a critical industrial system based on logs and metrics 1. https://doi.org/10.1109/EDCC.2018.00033
33. Janakiraman VM, Nielsen D (2016) Anomaly detection in aviation data using extreme learning machines
34. Lu H, Li Y, Mu S, Wang D, Kim H, Serikawa S (2017) Motor anomaly detection for unmanned aerial vehicles using reinforcement learning 4662(c). https://doi.org/10.1109/JIOT.2017.2737479
35. Bose B, Dutta J, Ghosh S, Pramanick P, Roy S (2018) D & RSense: detection of driving patterns and road anomalies. In: 2018 3rd International conference on internet of things smart innovation usages, pp 1–7. https://doi.org/10.1109/IoT-SIU.2018.8519861
36. Gibert X, Patel VM, Chellappa R (2016) Deep multitask learning for railway track inspection 1–12
37. Luo H, Zhong S (2017) Gas turbine engine gas path anomaly detection using deep learning with Gaussian distribution
38. Kang X, Song B, Sun F (2019) A deep similarity metric method based on incomplete data for traffic anomaly detection in IoT. Appl Sci 9(1):135. https://doi.org/10.3390/app9010135
39. Islam RU, Hossain MS, Andersson K (2018) A novel anomaly detection algorithm for sensor data under. Soft Comput 22(5):1623–1639. https://doi.org/10.1007/s00500-016-2425-2
40. Sun Y, Wu TY, Li X, Guizani M (2016) A rule verification system for smart building 1–12. https://doi.org/10.1109/TETC.2016.2531288
41. Kumar D, Bezdek JC, Rajasegarar S, Palaniswami M, Leckie C, Chan J (2016) Adaptive cluster tendency visualization and anomaly detection for streaming data 11(2)
42. Chen L, Member S, Ho Y, Hsieh H, Huang S, Lee H (2018) ADF: an Anomaly detection framework for large-scale PM2.5 sensing systems 4662(c):1–12. https://doi.org/10.1109/JIOT.2017.2766085
43. Antonini M, Vecchio M, Antonelli F, Ducange P, Perera C (2018) Smart audio sensors in the internet of things edge for anomaly detection. IEEE Access 6:67594–67610. https://doi.org/10.1109/ACCESS.2018.2877523
44. Manashty A, Light J, John S (2015) Healthcare event aggregation lab (HEAL), a knowledge sharing platform for anomaly detection and prediction 648–652
45. Lyu L, Jin J, Rajasegarar S, He X, Palaniswami M (2017) Fog-empowered anomaly detection in IoT using hyperellipsoidal clustering. IEEE Internet Things J 4(5):1174–1184. https://doi.org/10.1109/JIOT.2017.2709942
46. Rodriguez MA, Kotagiri R, Buyya R (2018) Detecting performance anomalies in scientific workflows using hierarchical temporal memory. Futur Gener Comput Syst 88:624–635. https://doi.org/10.1016/j.future.2018.05.014

47. Azimi I et al (2019) Personalized maternal sleep quality assessment: an objective IoT-based longitudinal study. IEEE Access 7:93433–93447. https://doi.org/10.1109/access.2019.2927781
48. Van Kasteren TLM, Englebienne G, Kröse BJA (2010) Activity recognition using semi-markov models on real world smart home datasets. J Ambient Intell Smart Environ 2(3):311–325
49. Legrand A, Niepceron B, Cournier A, Trannois H (2019) Study of autoencoder neural networks for anomaly detection in connected buildings. In: 2018 IEEE global conference on internet of things, GCIoT 2018, pp 1–5. https://doi.org/10.1109/GCIoT.2018.8620158

Gaze Prediction Based on Convolutional Neural Network

Abdulkader Helwan, Mohammad Khaleel Sallam Ma'aitah, Selin Uzelaltinbulat, Mohamad Ziad Altobel, and Manal Darwish

Abstract In this paper gaze prediction aims to acquire good performance via "Convolutional Neural Network (CNN)" based identification. Gaze prediction was proposed to localize the gaze direction of the students in classrooms using captured face images. Such systems can raise the teachers' awareness of whether students are gazing towards him or not and alert the students to pay attention to their teachers if their gazes were scattered. In this way, the prediction of gaze was proposed firstly to achieve the proposed eyes area detection algorithm based on modeling facial gestures. The inputs of the gaze prediction are the output of the eye detection. Secondly, the localizing gaze of the students had been considered as straight, right, or left. CNN was employed to different training models for the gaze prediction phase. The dataset was obtained from a public database for neural network training.

Keywords Gaze prediction · Eyes detection · Image capturing · CNN

A. Helwan
School of Engineering, Lebanese American University, Byblos, Lebanon

M. K. S. Ma'aitah
Department of Management Information Systems, Near East University, P.O.BOX: 99138, Nicosia, TRNC, Mersin 10, Turkey

S. Uzelaltinbulat (✉)
Department of Computer Information Systems, Near East University, P.O.BOX: 99138, Nicosia, TRNC, Mersin 10, Turkey
e-mail: selin.uzelaltinbulat@neu.edu.tr

M. Z. Altobel · M. Darwish
Department of Computer Engineering, Near East University, Nicosia, TRNC, Mersin 10, Turkey

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_18

1 Introduction

Recently, accurate image classification has been achieved by deep learning-based systems. Those deep networks showed superhuman accuracies in performing such tasks and this success motivated the researchers to apply those images for classification tasks. The results showed that deep networks can efficiently extract useful features that distinguish different image classes. Convolutional neural networks have been applied to various types of image diagnosis and classification due to their power of extracting different level features from images [1–3].

CNN began to gain popularity after being used in the AlexNet algorithm, which won the "ImageNet Large Scale Visual Recognition Challenge (ILSVRC)" competition. Reaching an accuracy close to human perception in restricted areas with the image recognition performance, the models trained with CNN have started to take place in many areas of our daily life. In convolutional neural networks, there are model parameters that might be learned from the pattern through training in the standard training process, as well as parameters that might not be learned from the data and must be defined by the designer in advance.

Many problems have encountered images or video capturing that needs to be processed with smart algorithms in everyday life. It is significant to find solutions to problems based on video and photographic images, as it guides efforts to increase the quality of life of the society [1]. In other words, the classification process with images, videos and similar data produced from satellites is performed using convolutional neural networks with high accuracy [2].

Computers can be used efficiently with gaze prediction. Moreover, gaze localization systems have found many applications in some areas as like eye-tracking systems, script reading systems, automotive design systems. Gaze prediction algorithm allows students to estimate the coordinates and angular distance of students' gazes on any target [3–7].

In this paper, the prediction of gaze direction of students had been indicated a similar and very useful task, if the student is gazing straight, left, or right. First, it localizes the direction of appearance of a student before perform regression. Such two-steps view prediction framework can significantly aid in improving the overview forecast performance, since regression modeling of the perspective forecast can be performed within a predefined range based on the predicted or classified directions. Neural networks (NNs) had been employed in many applications in order to learn algorithms and trained using public datasets [8–10]. Such NN algorithms have been employed in many machine learning areas, hence it had also been developed based on averaging output probabilities criterion including overall output aggregation [11–13]. Experimental results show the effectiveness of employing CNN under different training algorithms of the prediction of gaze direction [14].

2 Research Area

The main purpose of this research is to maintain an appearance based on gaze prediction system from captured face images of students in classrooms. Firstly, the face patterns had been processed purposes of the detected eyes regions. The images had been obtained from the Georgia Tech face images database [15]. It had been collected from the database which gaze patterns for individual looking 'straight', 'left', and 'right'.

3 Eye Detection and Gaze Prediction

It is important to detect the eyes areas in given pattern [15]. The faces are centered in the images and it has been described at the following subsection.

3.1 Image Processing

The original-colored facial images have been obtained and collected from students' gazing in different directions from Georgia Tech face database shown in Fig. 1. The sequence of the image processing illustrated in Fig. 2.

The colored patterns had been transformed to grayscale using the average model. After that, all images have been resized from their original size then presented to the eye detection phase.

3.2 Face Modeling for Eye Detection

The morphological structure of facial patterns suggest that one can effectively represent where the eyes regions lie. The left eye area has been located towards the

Fig. 1 Face images from the Georgia Tech face database [15]

Fig. 2 Image processing

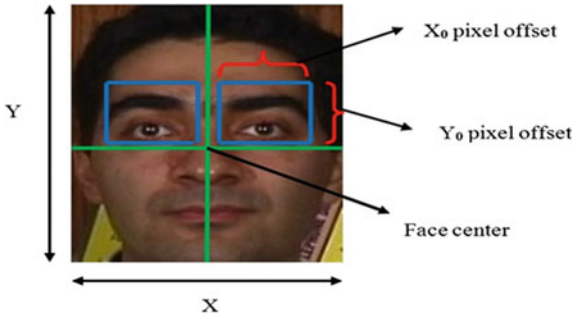


Fig. 3 Modeling for detection of the eye regions

left, spanning some pixels range which refer to within this work as X0 offset up from the nose center point. At the same time, while the right eye is located towards the right, spanning some pixels range which we refer to as Y0 offset. It has been taking the dimension or size of the image as X × Y, so that middle point of the image and therefore nose middle point have the pixel coordinate (X/2, Y/2). Figure 3 illustrates the modeling for detection of the eye regions.

The CNN has been received the output of the eye's detection phase, and then decided whether the students' appearance is to straight, right, or left. Considering the gaze prediction phase, the data sets which present eye patterns for training the NNs has been used.

4 Experiments, Results and Discussion

4.1 Eye Region Detection

CNN has been used to guide the direct mapping of extracted eyes region to the three gaze classes. Thus, it has been obtained from the segmented eye patterns from the public database and it covers the three gaze classes within this work.

Figure 4 has been illustrated for samples of unprocessed eyes images. Specifically, 200 images per gaze direction had been collected from totally 600 images. The obtained images had been reverted to grayscale patterns and then rescaled again to 32 × 32 (1024) pixels. The pixels coordinate for the nose center point is (60, 100) since the images supplied to the detection phase are of size 120 × 200 pixels. Particularly, from the processed 120 × 200 pixels face images, we extract 35–75 and 30–90 pixels along the vertical and horizontal axes

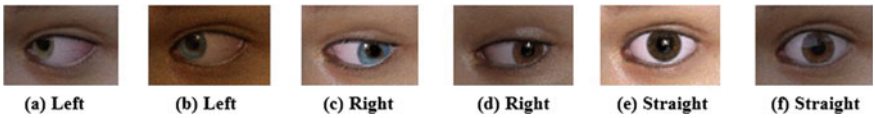


Fig. 4 Samples of different gazes [16]

respectively for the left eye detection. Similarly, we extracted 35–75 and 120–180 pixels along the vertical and horizontal axes respectively for the left eye detection.

Several experiments demonstrate the effectiveness of this simple approach for eyes regions detection given that a face has been already detected and duly registered.

Figure 5a, b illustrate the detected eyes regions for the student (Figs. 6 and 7).

Figure 8 illustrated that the CNN architecture where convolution layer represented by "Conv", batch normalization represented by "BN", feature maps represented by "FM" and fully connected layer represented by "FC" for classification.

In this paper, all convolution processes had been achieved using convolution filters of size (3 × 3) with zero padding and all pooling processes had been carried out by using max pooling windows of size (2 × 2) and lastly the input images to the model are of size (32 × 32) [17–23].

For optimizing the proposed model, mini-batch optimization via gradient descent was used as a batch size of 60 [24–29]. |Additionally, the learning rate was taken as 0.001 and trained the model for 100 epochs. The "learning curve (LC)" for the trained convolutional neural network is illustrated in Fig. 9, with validation accuracy of 90.67%.

In this research, the obtained validation accuracy is 90.76%. The proposed algorithm has been tested using facial images obtained for eyes detection and gaze prediction from a different database, Georgia Tech face database [15]. Figure 10 illustrates the collected face images with different gaze directions for testing the aforementioned database.

Experimental results show the classification of the gaze direction with an accuracy of 88.33%.

Fig. 5 Proposed framework

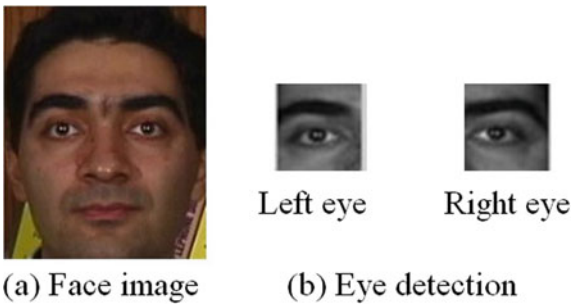


Fig. 6 Training images

Fig. 7 Testing images

Moreover, that proposed approach is only used for the synthetic dataset and tested by using real images. This can alleviate the problem of collecting real images for developing eye gaze prediction systems.

5 Comparison with Related Works

Studies that focused on detecting gaze directly from eyes gazing directions are quite a few. In a study by Lu et al. [30] an extraction of the 8D or 15D intensity features from the eye region was carried out. The authors used subpixel alignment regression (SAR) for gaze estimation, and they reported an accuracy of 62%.

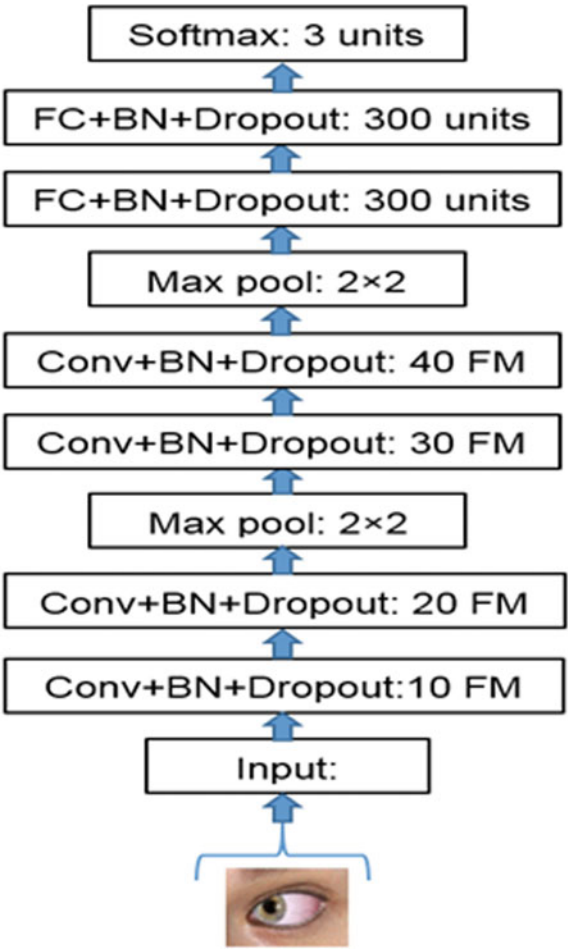


Fig. 8 Proposed CNN architecture

Moreover, George [31] proposed an eye tracking system using convolutional neural network. The work presents a new framework for biometric identification based on eye movement parameters. It was stated that such a system was capable of achieving an accuracy of 85.65% during its testing phases.

	SAR [30]	CNN [31]	Ours
Testing accuracy (%)	62	85.65	87.33

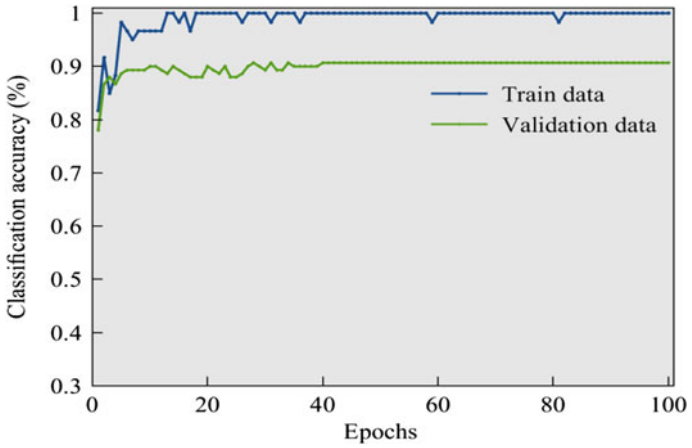


Fig. 9 Trained CNN with LC

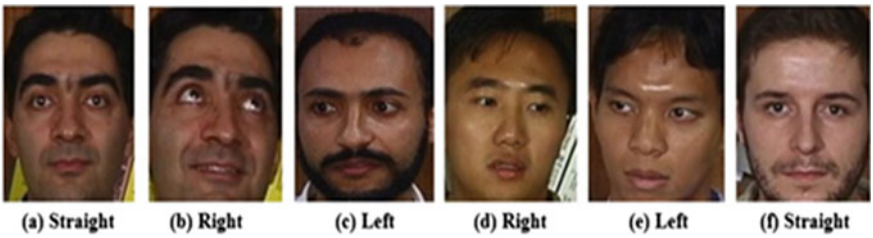


Fig. 10 Detected gaze images during the testing phase

6 Conclusions

The paper is devoted to the synthesis of NN based gaze prediction system. The system includes pre-processing and classification stages. These features are input signal for NN. NN is applied to classify the gaze patterns in a prediction step. The operation principle and learning algorithm of NN based gaze prediction system are presented. Gaze prediction permits to computers to indicate students' gaze directions.

Firstly, it is assumed that face has been centrally registered. Secondly, it has been implemented to a unique facial pattern to carry out the gaze prediction. We train the designed CNN on the segmented eye images and obtain a promising testing accuracy of 88.33%. Hence, the eyes detection module might not capture main eyes regions in failing registered face patterns. In order to improve the efficiency of the phase detection module, "Support Vector Machine (SVM)" might be used to find out what the eyes detection module capture.

References

1. Jian-Gang W, Eric S, Ronda V (2003) Eye gaze prediction from a single image of one eye. In: 9th International conference on computer vision, France, pp 136–143
2. Jixu C, Qiang J (2011) Probabilistic gaze prediction without active personal calibration. In: Proceedings of international conference of computer vision and patterns recognition, Colorado Springs, CO, pp 609–616
3. Feng L, Yusuke S, Takahiro O, Yoichi S (2014) Adaptive linear regression for appearance-based gaze prediction. Trans Pattern Anal Mach Intell 36(10):2033–2046
4. Xiao-Meng W, Kang L, Xu Q (2015) A survey on gaze prediction. In: Proceedings of international conference on intelligent systems and knowledge engineering, pp 260–267
5. Amer A, Miad F (2013) Eye tracking and head movement detection: a state of-art survey. IEEE J Trans Eng Health Med 1
6. Sireesha M, Vijaya P, Chellamma K (2013) A survey on gaze prediction techniques. In: Proceedings of international conference on VLSI communication advanced devices signals and systems and networking, pp 353–361
7. Xucong Z, Yusuke S, Mario F, Andreas B (2015) Appearance based gaze prediction in the wild. In: Conference on computer vision and pattern recognition (CVPR), pp 4511–4520
8. Bush IJ, Abiyev RH, Mohammad KM (2017) Intelligent machine learning algorithms for colour segmentation. WSEAS Trans Signal Process 13:232–240
9. Abiyev RH, Helwan A (2018) Fuzzy neural networks for identification of breast cancer using images' shape and texture features. J Med Imag Health Inf 8(4):817–825
10. Abdulkader H, Adnan K, Ebenezer O, Oyebade O, Oyetade O (2016) Seminal quality evaluation with RBF neural network. Bulletin of the Transilvania University of Braşov. No 2, Series III: Math Inf Phy 9(58):137–146
11. Helwan A, Menekay M, Ma'aitah MKS (2020) Machine learning for better understanding of autistics. In: International conference on theory and applications of fuzzy systems and soft computing, pp 817–825. Springer, Cham
12. Helwan A, Ma'aitah MKS (2020) Machine learning in automated chest radiographs classification. In: International conference on theory and applications of fuzzy systems and soft computing, pp 810–816. Springer, Cham
13. Oyebade O, Kamil D (2016) Pattern recognition: invariance learning in convolutional auto encoder network. Int J Image, Graph Signal Process 8(3):19–27
14. Kyung-Nam K, Ramakrishna R (1999) Vision-based eye-gaze tracking for human computer interface. In: IEEE SMC'99 conference proceedings. 1999 IEEE international conference on systems, man, and cybernetics, Tokyo, Japan, pp 324–329
15. Georgia Tech Face database (2016) Retrieved on 2 Aug 2016. Available at: http://www.anefian.com/research/face_reco.htm
16. Erroll W, Tadas B, Xucong Z, Yusuke S, Peter R, Andreas B (2015) Rendering of eyes for eye-shape registration and gaze prediction. In: IEEE international conference on computer vision (ICCV), Santiago, Chile, pp 3756–3764
17. Erroll W, Tadas B, Xucong Z, Yusuke S, Peter R, Andreas B (2015) Rendering of eyes for eye-shape registration and gaze prediction. In: International conference on computer vision (ICCV), Santiago, Chile, pp 3756–3764
18. Kyle K, Aditya K, Petr K, Harini K, Suchendra B, Wojciech M, Antonio T (2016) Eye tracking for everyone. In: The IEEE conference on computer vision and pattern recognition (CVPR)
19. Chennamma H, Xiaohui Y (2013) A survey on eye-gaze tracking techniques. Indian J Comput Sci Eng 4(5):388–393
20. Matthew Z, Rob F (2013) Stochastic pooling for regularization of deep convolutional neural networks. In: 1st International conference on learning representations, Scottsdale, United States, arXiv preprint arXiv:1301.3557

21. Hoo-Chang S, Holger R, Mingchen G, Le L, Ziyue X, Isabella N, Jianhua Y, Daniel M, Ronald S (2016) Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. IEEE Trans Med Imaging 35 (5):1285–1298
22. Dumitru E, Yoshua B, Aaron C, Pierre-Antoine M, Pascal V (2010) Why does unsupervised pre-training help deep learning? J Mach Learn Res 11:625–660
23. Ossama A, Abdel-rahman M, Hui J, Gerald P (2012) Applying convolutional neural networks concepts to hybrid NN-HMM model for speech recognition. In: Proceedings of the IEEE international conference on acoustics, speech, and signal processing (ICASSP '12), pp 4277–4280, IEEE, Kyoto, Japan
24. Nitish S, Geoffrey H, Alex K, Ilya S, Ruslan S (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(56):1929–1958
25. Sergey I, Christian S (2015) Batch normalization: Accelerating deep network training by reducing internal covariate shift. Int Conf Mach Learn 37:448–456
26. Kaiming H, Jian S (2015) Convolutional neural networks at constrained time cost. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR '15), Boston, Mass, USA, pp 5353–5360
27. Steve L, Lee G, Ahchung T, Andrew B (1997) Face recognition: a convolutional neural-network approach. IEEE Trans Neural Networks 8(1):98–113
28. Yaniv T, Ming Y, Marcaurelio R, Lior W (2014) DeepFace: closing the gap to human-level performance in face verification. In: Conference on computer vision and pattern recognition (CVPR), pp 1701–1708
29. Abiyev RH, Ma'aitah MKS (2018) Deep convolutional neural networks for chest diseases detection. J. Healthc Eng
30. Lu F, Sugano Y, Okabe T, Sato Y (2011) Inferring human gaze from appearance via adaptive linear regression. In: Metaxas DN, Quan L, Sanfeliu A, Gool LJV (eds) Proceedings of the IEEE international conference on computer vision (ICCV '11), pp 153–160, IEEE, Barcelona, Spain, November
31. George A (2019) Image based eye gaze tracking and its applications, arXiv preprint arXiv:1907.04325

Stacked Autoencoders Deep Learning Approach for Left Ventricular Localization in Magnetic Resonance Slices

Abdulkader Helwan, Mohammad Khaleel Sallam Ma'aitah, Selin Uzelaltinbulat, Bengi Sonyel, Mohamad Ziad Ziad Altobel, and Manal Darwish

Abstract Deep learning (DL) is an effective method for medical object detection. Studies show that deep networks can achieve accuracy in medical segmentation and detection tasks. This is due to the depth and training methods of deep networks which allows them to derive different levels of abstractions of input mages. In this paper, the left ventricle detection task is carried out using a deep network called stacked auto-encoder (SAE). The networks take off this task as a binary classification task wherein left and non-left ventricles cropped images are being recognized by the SAE. Once the network recognizes left and non-left ventricles, the whole task starts by initiating a sliding window that moves through the whole magnetic resonance (MR) slice till a left ventricle is detected. Experimentally, the network showed effective detection performance when target images are noisy as it is seen that it can detect left ventricles in target images with up to 10% of salt and pepper noise.

A. Helwan
School of Engineering, Lebanese American University, Byblos, Lebanon

M. K. S. Ma'aitah (✉)
Department of Management Information Systems, Near East University, Mersin 10, Nicosia, TRNC, Turkey
e-mail: mohammad.maaitah@neu.edu.tr

S. Uzelaltinbulat
Department of Computer Information Systems, Near East University, Mersin 10, Nicosia, TRNC, Turkey

B. Sonyel
Department of Educational Sciences, Eastern Mediterranean University, Mersin 10, Famagusta, TRNC, Turkey

M. Z. Z. Altobel · M. Darwish
Department of Computer Engineering, Near East University, Mersin 10, Nicosia, TRNC, Turkey

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_19

Keywords Sliding windows · Deep learning · Stacked auto-encoder · Binary classification · Intelligent system

1 Introduction

Deep networks have seen a rapid rise in different areas in particular medicine, to solve different conditions in various medical areas: tumor diagnosis [1, 2], image classification [3, 4], and image segmentation [5]. The high efficiency obtained by using the deep networks in medicine motivated researchers to apply these deep models to localize medical objects in a similar and more accurate manner to humans. Therefore, many studies have used deep networks to detect and localize different organs in medical images such as Left and right ventricles in the MR images [5, 6].

The segmentation and detection of the body organs, in particular the heart, is still a tedious task for radiologists. Therefore, the development of computer systems that helps radiologists in interpreting and detecting the body's sophisticated organs is in immense need. Recently, deep neural networks (DNNs) have been extensively applied to solve various medical problems [6–11]. DL refers to machine learning (ML) models that have a deep structure. It grants a great capability in obtaining the medium and high-level abstractions from raw input data [3].

Recently, DNNs, particularly convolutional neural networks (CNNs) and auto-encoders (AEs) have received a great deal of attention from researchers in the medical field, due to their great activity in image classification and detection [1–4]. It is notable that the depth and pre-training methods of deep networks allow them to score higher accuracies than those of conventional shallow networks [12–14].

Thus, in this paper, we opt to design a deep approach for the localization of left ventricles in MR slices, obtained from the Sunny Brook database [7]. The system is based on a deep network named SAE, which is trained to scan an MR slice to localize the left ventricle. Note that this work is an extended version of an earlier published work on the localization of left ventricles using a conventional back-propagation neural network (BPNN) [8]. Hence, in this work, we aim to apply a deep network (SAE) instead of BPNN and investigate the differences in terms of accuracy, training time and mean square error achieved.

Several researches have been conducted on the segmentation and localization of ventricles. In [9] the authors have directed an examination on the utilization of a Fully Convolutional Neural Network (FCN) for Cardiac Segmentation in Short-Axis MRI. The proposed FCN design utilized was effectively prepared start to finish on a graphic processing unit (GPU) in a solitary learning stage, to make derivation at each pixel. This model was validated, and it freely fragments each picture in milliseconds, along these lines it could be utilized in parallel on bunches of CPUs, GPUs, or both for adaptable and accurate ventricle segmentation.

In another research [10], the authors trained a CNN with several pre-processing and data augmentation steps aimed at generalizing and preventing overfitting using

image filters. Lastly, they were able to demonstrate that the left ventricle can be detected as a single-filled bordering region.

This paper designed as follows: section one is an introduction of the work, in which aims, and literature are discussed. Section two presents a brief explanation of the proposed approach, in which AEs are discussed. Section three is a discussion of the results obtained, in addition to the comparison of the performances of the proposed system with other related works. Finally, section four is the conclusion of the work presented.

2 Materials and Methods

2.1 Left Ventricle Recognition

In this section, a SAE is prepared to perceive left ventricle and non-left ventricle detection. Hence, to accomplish this we utilize a paired order task, i.e., we utilized left ventricles and non-left ventricle images to prepare the network. Data used for validating the network are collected from Sunnybrook Cardiac Data (SCD) [7]. We chose 3000 images from the database and utilized them to get cropped left and non-left ventricles. Along these lines, left ventricles were made by editing the MR slices to just have left ventricles and are alluded to as positive models. Then again, the non-left ventricle pictures are made by editing the MR slices to have some other irregular shapes or articles barring the left ventricle and are alluded to as negative models. This was considered, as it appears to improve the power of the framework in perceiving left ventricle and non-left ventricle pictures. Trimming ventricle and non-ventricle images from heart MRI has appeared in Fig. 1.

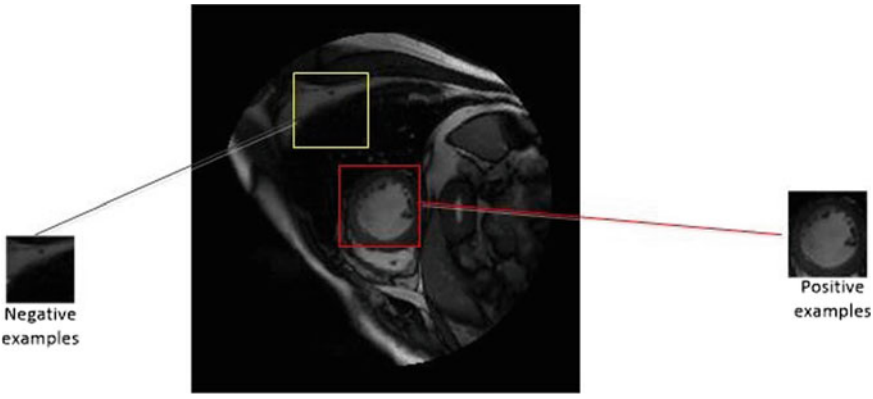


Fig. 1 Left and non-left ventricles extraction

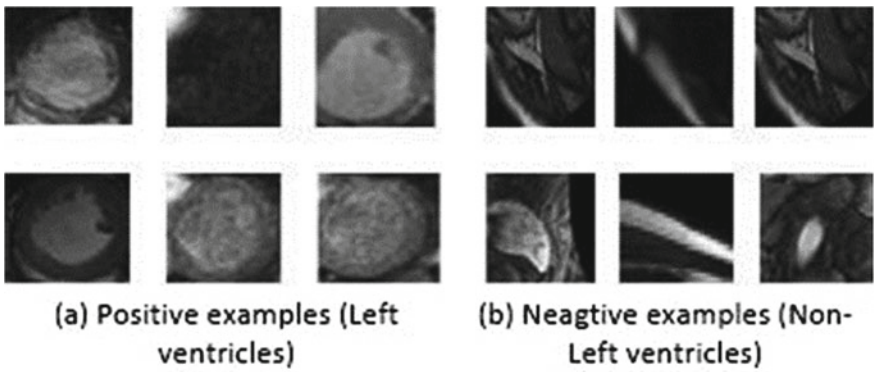


Fig. 2 Example of positive and negative images

Note that trimmed images are resized to 40 × 40 pixels (1600 pixels) so as to make the pictures reliable. Tests of positive and negative models have appeared in Fig. 2.

2.2 Stacked Auto Encoder: Design, Training, and Testing

A SAE is a neural network that mainly consists of two or more hidden layers that are stacked together after pre-training using an algorithm called Greedy Layer-Wise Training. This algorithm is a method used to train separate layers of AEs in an unsupervised manner. After pre-training, the network layers are stacked together and fine-tuned using the backpropagation algorithm.

The network is trained and simulated on a dual-core, intel (R) Core (TM) (2.6 GHz) CPU with 4 GB RAM, with programs written and run in a MATLAB environment.

A SAE of two hidden layers utilized to learn images of size 40 × 40 (1600 pixels). The 300 images were partitioned into training and validation data where the validation data permits the perception of execution of the prepared SAE on inconspicuous or new data. 200 examples containing both left and non-left ventricles were used for training the network, whereas the rest were used for testing. Figure 3 shows the network architecture in which 2 hidden layers are used, as two AEs are stacked to create this network.

Similar pictures were utilized for tweaking this network (AE) which comprises 1600 neurons since each image is rescaled to 40 × 40 pixels. The two hidden layers comprised of 85 and 50 neurons, individually; while the output layer has 2 neurons since the 2 classes, we have only two: left ventricle or non-left ventricle (Fig. 3).

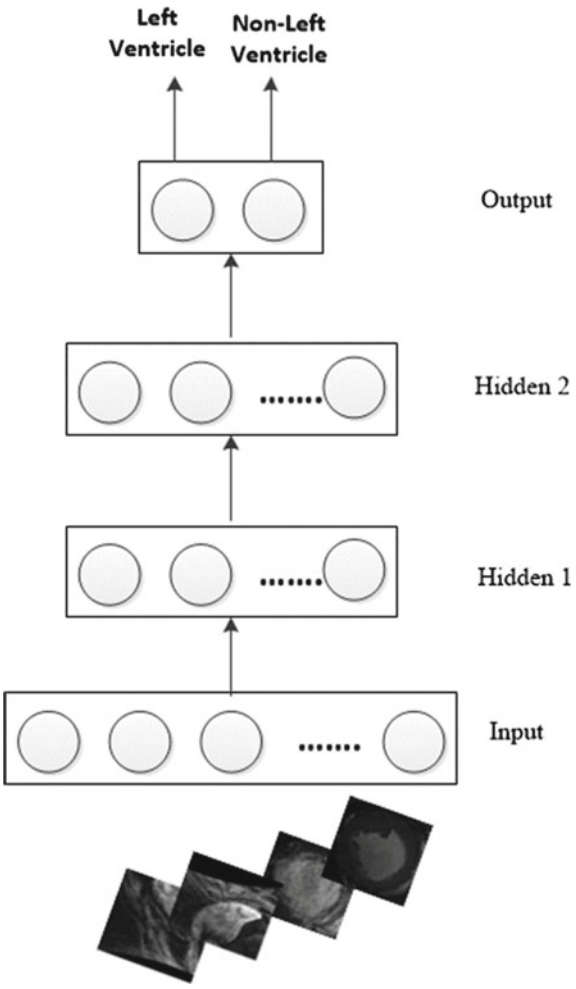


Fig. 3 SAE training on positive and negative examples

Table 1 shows the input parameters setting of the system that was utilized in training the system. The network kept running for 200 most extreme cycles and achieved a mean square error of 0.012 and a 100% recognition rate during pre-training in 40 s. In any case, the time was longer during the calibrating where the network achieved a mean square error of 0.025 and 98% recognition rate in 9 s.

Table 2 shows the training and testing recognition rates achieved by the SAE when trained and tested 200 and 100 left and non-left ventricles, respectively.

Table 1 Final training parameters

Number of training samples	300
Number of neurons	1600
Number of hidden neurons	85–50
Activation function at hidden and output layers	Log-Sigmoid
Learning rate (η)	0.21
Momentum rate (α)	0.71
Minimum square error (MSE)	0.012
Epochs	200
Training time	40 s

Table 2 SAE training and testing performance

Parameter	Training	Testing
Number of samples	200	100
Number of samples correct classified	200	95
Recognition rates (%)	100	95

2.3 *Left Ventricle Detection in MR Slice*

After training the network to recognize left and non-left ventricles it is now time to evaluate it on a complete MR slice containing various objects and find out if it can localize the left ventricle. Hence, as to distinguish left ventricles in new images, the new target images are examined in a non-covering design utilizing a sliding window or veil as appeared in Fig. 4. As seen, images converted to 120 * 120 pixels while the sliding window is set to be of 40 * 40 pixels. This allows the window to fit in without tumbling off the image's edges.

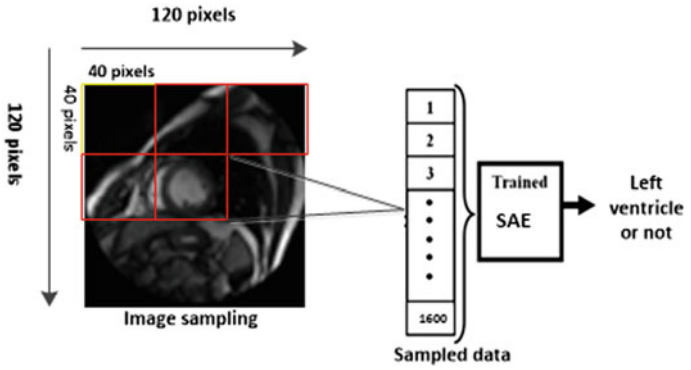


Fig. 4 SAE training on positive and negative examples

3 Results and Discussion

For the purpose of showing the adequacy of the developed framework, we evaluated tests utilizing objective images gathered from the database. Also, we encountered the performance of the system by adding commotion to the objective images. The idea is to include salt and pepper noise at various commotion levels to the first target images and examine the capacity of the prepared SAE in confining the left ventricle. This additionally means to investigate at what noise level; the developed framework begins to recognize left and non-left ventricles in the objective images. Figures 5 and 6 show some examples of correct and incorrect left ventricle detection.

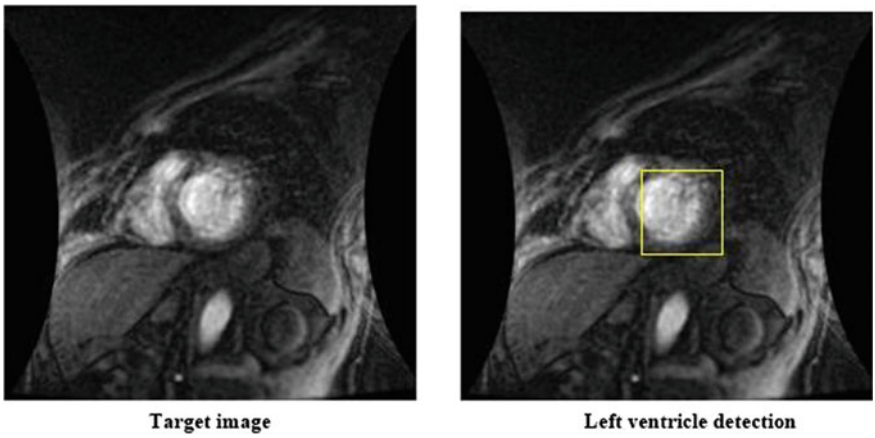


Fig. 5 Detection outcome using the developed system

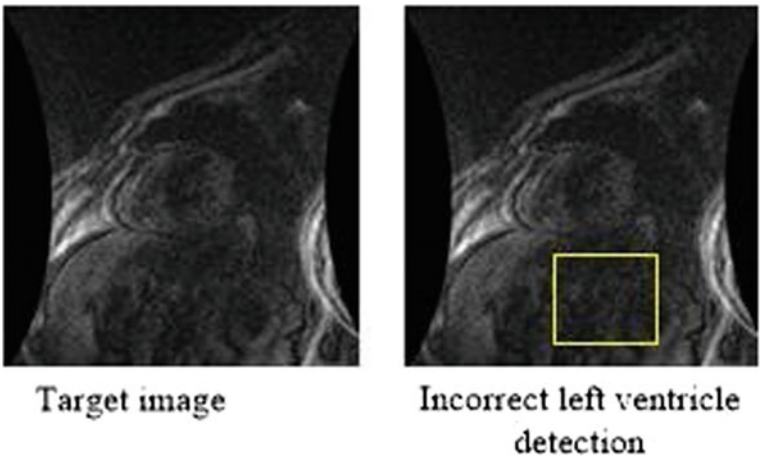


Fig. 6 Incorrect detection of the left ventricle

Figures 7, 8 and 9 shows the detection outcomes after adding noise to some target images.

For the first run, the system failed to localize the left ventricle of the image that contains 10% of salt and pepper noise (Fig. 8). However, one more run was enough for the system to learn and be capable of detecting the left ventricle in the same image. This is shown in Fig. 9.

Table 3 shows developed system performance with other earlier works. It is remarkable that this DL-based system which uses a SAE outperformed the other system [6], which uses a conventional backpropagation neural network. Note that this outperformance is in terms of generalization power (accuracy), time, and error rate.

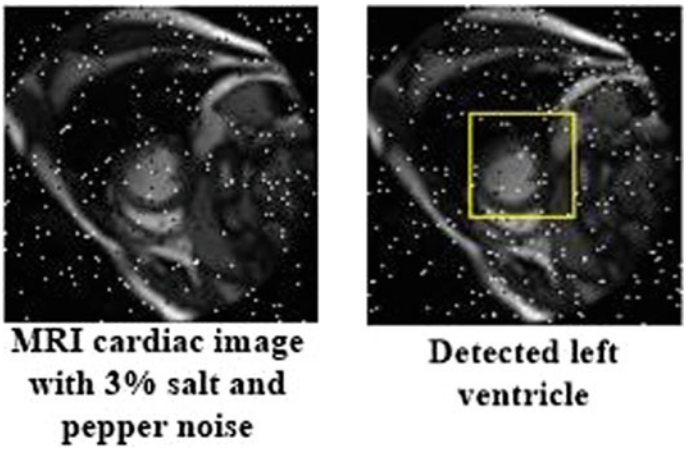


Fig. 7 Results of correct detection of the left ventricle of images with 3% noise

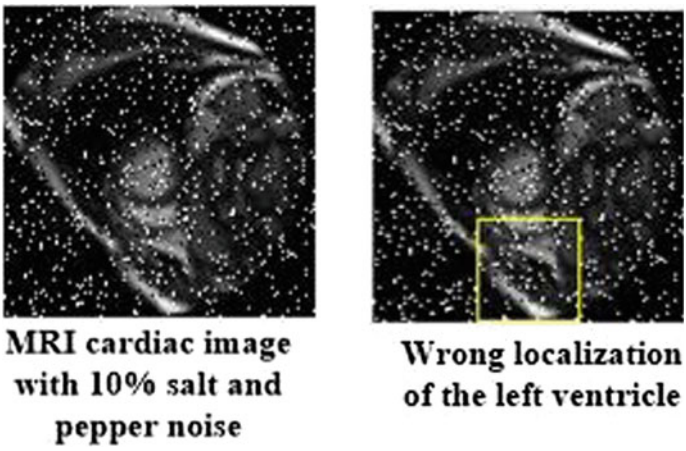


Fig. 8 Results of wrong detection of the left ventricle of images with 10% noise

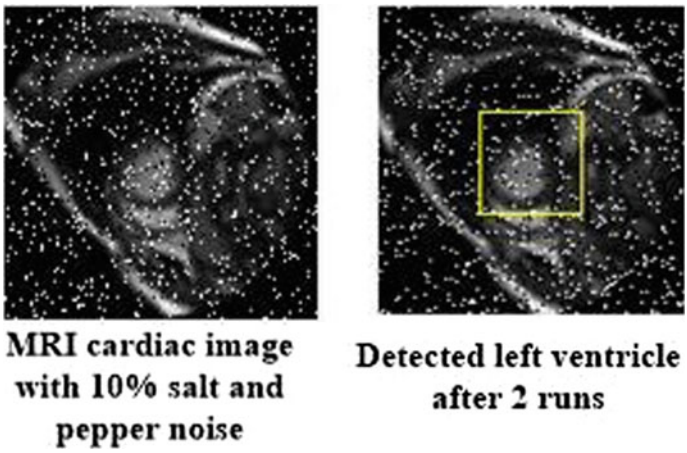


Fig. 9 Results of detection of the left ventricle of images with 10% noise after 2 runs

Table 3 Performance comparison

	SAE	BPNN [6]
Training recognition rate (%)	100	100
Testing recognition rate (%)	95	88
MSE achieved	0.012	0.028
Maximum number of iterations	200	1215
Training time	52 s	40

4 Conclusion

In this work, a deep network named SAE has been utilized as the learning framework in this exploration. The SAE is prepared on tests of left ventricles and non-left ventricles (irregular) gathered from the MR slices. For the identification of left ventricles in target images, a window estimate (40 × 40 pixels) compared to the size of the input to the SAE, is utilized to test the SAE on objective images in a non-covering design. The network is validated using some arbitrarily gathered target images containing left ventricles with no scene limitations, i.e., the scale, interpretation, brightening, direction, and so forth of left ventricles contained in the objective images. The developed framework apparently performs very well in identifying left ventricles, including situations where left ventricles are even incompletely blocked. Besides, to demonstrate the adequacy and power of the designed system, it was tested on uproarious and noisy target images.

Finally, the performance of the deep network used in this research was compared to the one achieved by our earlier published work. Overall, it is seen that a deep network can outperform a conventional network in localizing the left ventricles in MR slices.

References

1. Oyedotun OK, Olaniyi EO, Helwan A, Khashman A (2015) Hybrid auto encoder network for iris nevus diagnosis considering potential malignancy. Proceedings of the international conference on advances in biomedical engineering (ICABME'15) (2015), pp 274–277
2. Abiyev RH, Helwan A (2018) Fuzzy neural networks for identification of breast cancer using images' shape and texture features. J Med Imag Health Inf 8(4):817–825
3. Abiyev RH, Ma'aitah MKS (2018) Deep convolutional neural networks for chest diseases detection. J Healthc Eng
4. Helwan A, Abiyev RH (2015) ISIBC: an intelligent system for identification of breast cancer. Proceedings of international conference on advances in biomedical engineering (ICABME'15), pp 17–20
5. Huang S, Liu J, Lee L et al (2009) Segmentation of the left ventricle from cine MR images using a comprehensive approach. Proceedings of the MICCAI 2009 workshop on cardiac MR left ventricle segmentation Challenge, MIDAS J
6. Marak L, Cousty J, Najman L, Talbot H (2009) 4D morphological segmentation and the MICCAI LV-segmentation grand challenge. Proceedings of the MICCAI 2009 workshop on cardiac MR left ventricle segmentation challenge, MIDAS J 1–8
7. Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright GA (2009) Evaluation framework for algorithms segmenting short axis cardiac MRI. The MIDAS J Cardiac MR left ventricle segmentation challenge
8. Helwan A, Dilber UO (2017) Sliding window based machine learning system for the left ventricle localization in MR cardiac images. Appl Comput Intell Soft Comput (2017)
9. Tran PV (2016) A fully convolutional neural network for cardiac segmentation in short-axis MRI. arXiv preprint arXiv:1604.00494 (2016)
10. Zotti C, Luo Z, Humbert O, Lalande A, Jodoin PM (2017) GridNet with automatic shape prior registration for automatic MRI cardiac segmentation. arXiv preprint arXiv:1705.08943
11. Ma'aitah MKS, Abiyev RH, Bush IJ (2017) Intelligent classification of liver disorder using fuzzy neural system. Int J Adv Comput Sci 8(12)
12. Helwan A, Menekay M, Ma'aitah MKS (2020) Machine learning for better understanding of autistics. In: International conference on theory and applications of fuzzy systems and soft computing. Springer, Cham, pp 817–825
13. Helwan A, Ma'aitah MKS (2020) Machine learning in automated chest radiographs classification. In: International conference on theory and applications of fuzzy systems and soft computing. Springer, Cham, pp 810–816
14. Bush IJ, Abiyev RH, Mohammad KM (2017) Intelligent machine learning algorithms for colour segmentation. WSEAS Trans Sig Process 13:232–240

Machine Learning Based Prediction of Edematous Malnutrition in Afghan Children

Ziaullah Momand, Mohammad Shuaib Zarinkhail, and Mohammad Farhad Aryan

Abstract Proper nutrition is one of the necessities of human health. Malnutrition in children is an ecumenical problem that causes mortality and morbidity in children. Afghanistan is one of the countries which suffers from child severe acute malnutrition. In this study, a machine learning based model was proposed to predict the severity of edematous malnutrition in children between the age of 1–59 months in the context of Afghanistan. Random Forest, J48, and Naïve Bayes classifiers were applied to the malnutrition-related data, which was collected from two hospitals in Afghanistan. The Random Forest technique obtained outstanding results with the highest accuracy of 97.14% and the performance of J48 was also moderate with 94.51% accuracy. This study explores how machine learning classification techniques can classify the edematous malnutrition in under 5 (U5) children. Overall, the findings of the proposed method demonstrate that our model with a robust result provides a potential mechanism for prediction of nutritional oedema using machine learning classification algorithms for Afghan U5 children that helps policy makers for child's malnutrition.

Keywords Oedema · Edema · Edematous Malnutrition · Under five children · Malnutrition · Severe Acute Malnutrition

1 Introduction

An adequate amount of nutritious food is essential for individuals, especially in the aging group population. In early childhood, proper nutritious food intake helps in healthy growth, proper intellectual abilities, and a strong immune system in children. Poor nutrition in children causes physical and mental development [1]. The term malnutrition is used for all bad forms of nutrition for individuals and remains

Z. Momand (✉) · M. F. Aryan
Faculty of Computer Science, Kabul University, Kabul, Afghanistan

M. S. Zarinkhail
Faculty of Education, Takhar University, Taloqan, Afghanistan

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_20

one of the most underlying causes of mortality in U5 children in the globe. Nutritional oedema is one of the most dangerous forms of malnutrition that causes Severe Acute Malnutrition (SAM) in children [2, 3].

SAM is a global problem in U5 children. According to the WHO's report, 47 million U5 children are globally wasted and 14.3 million are severely wasted. Moreover, 144 million U5 children are infected from the stunted condition. Nearly 45% of deaths of the U5 children have been associated with malnutrition [4]. However, the most vulnerable countries are the low and middle-income countries. Among the low-income countries, Afghanistan is one of the countries that has been suffering from SAM among children under five years of age and pregnant and lactating women (PLW). Afghanistan has one of the highest ranks of stunted children in the globe. Nearly, 41% of children U5 are stunted. The rate of wasted children is also considerable with a statistic of 9.5% in this country. Most of these children belong to poor families, where no basic health facilities are available [5]. Afghanistan is experiencing the worst situation of malnutrition-related mortality in U5 children and PLW.

Some studies [6–8] have been carried out to analyze the determinants and associated risk factors of malnutrition in U5 children and PLW in Afghanistan. According to their findings, Afghanistan was in a critical situation for children's malnutrition-related mortality. From the technology perspectives, no attempt has been done to overcome this problem. Hence, applying the potentials of Artificial Intelligence (AI) for the determination of the child's edematous malnutrition is a demand for new research in this field. Thus, the machine learning predictive models will be highly helpful to identify the malnutrition-related risk factors and determinants to reduce malnutrition-related child mortality in Afghanistan. Hence, this study aimed to explore the Afghan children's nutritional data and build a model to identify edematous malnutrition in children with the age of 1–59 months using machine learning algorithms.

The rest of this paper is organized in 5 sections: Sect. 2, gives you the current assessment procedure of SAM in Afghanistan. Related work is outlined in Sect. 3, and the methodology is described in Sect. 4. Section 5 explains the result of work and Sect. 6 contains the discussion. Finally, our conclusion and future direction are drawn in the final section of the paper.

2 Severe Acute Malnutrition Assessment in Afghanistan

The Ministry of Public Health (MoPH) of the Islamic Republic of Afghanistan developed a national nutritional strategy to prevent child and PLW malnutrition-related mortality in the country. In this strategy, the MoPH developed a community-based growth monitoring and promotion (cGMP) program for Community Health Works (LCW) to help in providing consultation and assessment for the child's nutritional status [3]. Currently, the SAM for U5 children measures

in two ways: assessment of the child's nutritional status through anthropometric parameters and clinical signs, which is known as nutritional oedema.

2.1 Edematous Malnutrition

Oedema is the water reservation in body tissues and is a clinical symptom for malnutrition identification. It always appears in bilateral pitting in both feet that confirms SAM or wasted status in children [9]. Oedema can result in other kinds of diseases rather than malnutrition such as heart failure, kidney damage, liver disease, etc. [10]. However, nutritional oedema starts from both feet and based on its severity, it can be extended upward of the body. To confirm the presence of nutritional oedema, the nutritionist applies the thumbs pressure to both feet for three seconds. If the pit persists on both feet when the thumbs are lifted, then the child is considered as edematous malnourished [3].

Nutritional oedema based on its severity is classified into three major classes. Mild oedema grade (+) emerges in bilateral pitting of both feet, moderate oedema grade (++) appears in bilateral pitting of both feet and lower limbs/hands, and the severe oedema grade (+++) is a generalized nutritional oedema of the feet, lower limbs, and periorbital area of eyes. Grade3 nutritional oedema is the most dangerous form that may cause other diseases [3].

2.2 SAM Assessment Through Anthropometric Parameters

The measurement of body parameters is the prominent indicator of malnutrition in children between the age of 6–59 months and PLW. SAM in U5 children can be assessed by measuring the Mid-Upper Arm Circumference (MUAC) and weight for height indicators. The MUAC measurement is a common indicator of SAM in Afghanistan [3]. However, it is not recommended for infants less than six months of age. Table 1 indicates the MUAC measurement cut-offs for SAM identification.

Measurement of weight for age is another indicator of SAM in U5 children. Measuring the weight for age z-score (WHZ) from the median of WHO's growth velocity standard [3] indicates the severe weight gain or wasted status. Table 2 shows the WHZ cut-offs for the SAM identification based on the WHO growth standards.

Table 1 MUAC Measurement Cut-offs for SAM Identification

Normal Nutrition	Moderate Acute Malnutrition	Severe Acute Malnutrition
MUAC >= 125 mm	MUAC >= 115 mm and < 125 mm	MUAC < 115 mm

Table 2 WHZ measurement cut-offs for SAM identification

Normal Nutrition	Moderate Acute Malnutrition	Severe Acute Malnutrition
WHZ >= −2	WHZ >= −3 and < −2	WHZ < −3

3 Related Work

Applying machine learning algorithms to nutritional data opened a new door for researchers to reduce the child and PLW nutrition-related mortality and morbidity in the target population. Different machine learning algorithms were applied to explore the hidden pattern in various malnutrition-related data to help the policymakers to identify the nutritional status of U5 children and PLW as easily as possible. Recently, A Random Forest, PART rule, Naïve Bayes, and Logistic Regression classifiers were applied to the Afghanistan nutrition SMART survey data. The study was aimed to predict three malnutrition statuses in Afghan U5 children, which were stunted, underweight, and wasted. Furthermore, in this study, the presence of nutritional oedema was also investigated. The empirical results of the Random Forest classifier showed better performance rather than other classifiers [11].

A logistic regression model was built to predict the nutritional status of U5 children based on the anthropometric parameters using Indian Demographic and Health Survey (IDHS) data [12]. Similarly, Bangladesh DHS data was explored in the context of Bangladesh to predict anthropometric based nutritional status in children with the age of 1–59 months using logistic regression classifier [13].

Artificial Neural Network (ANN) is a supervised machine learning technique that handles various types of clinical data and provides useful output to the health industry decision-makers [14]. An ANN predictive model was built to predict the SAM of children based on the routine growth curve [15]. Likewise, a deep learning-based model was developed to predict the malnutrition status of U5 children based on their anthropometric parameters [16].

Moreover, K-nearest neighbor (KNN), support vector machine (SVM), random forest, and logistic regression classifier techniques were applied to the BDHS dataset to predict the malnutrition status among U5 children in the context of Bangladesh [17]. Likewise, in a case study in Sri Lanka, a decision tree classifier was applied to the Sri Lanka DHS dataset. The authors utilized trees to develop rules for malnutrition identification among under-five children [18]. Similarly, a number of useful rules for malnutrition detection in under-five children were generated from Ethiopian DHS data using PART rule induction, random forest, and naïve Bayes classifiers [19].

As pointed in the above previous related work, most of the studies developed predictive models based on anthropometric features and the child's food intakes. However, this study aimed to classify nutritional oedema based on its severity using machine learning algorithms that have not been done before.

4 Methodology

Machine learning approaches always try to extract the hidden pattern of historical data in an organized and systematic way. In this study, the Afghan children's nutritional oedema data was collected and investigated for building a model to classify the severity of edematous malnutrition. This work was carried out in an organized manner as depicted in Fig. 1.

4.1 Data Source

For this study, a secondary data was collected from the two hospitals in Afghanistan, which are the provincial hospital of Parwan province and Baraki-Rajan hospital in Logar province. The data was collected from the databases of routine growth monitoring and malnutrition screening of U5 children in the hospitals.

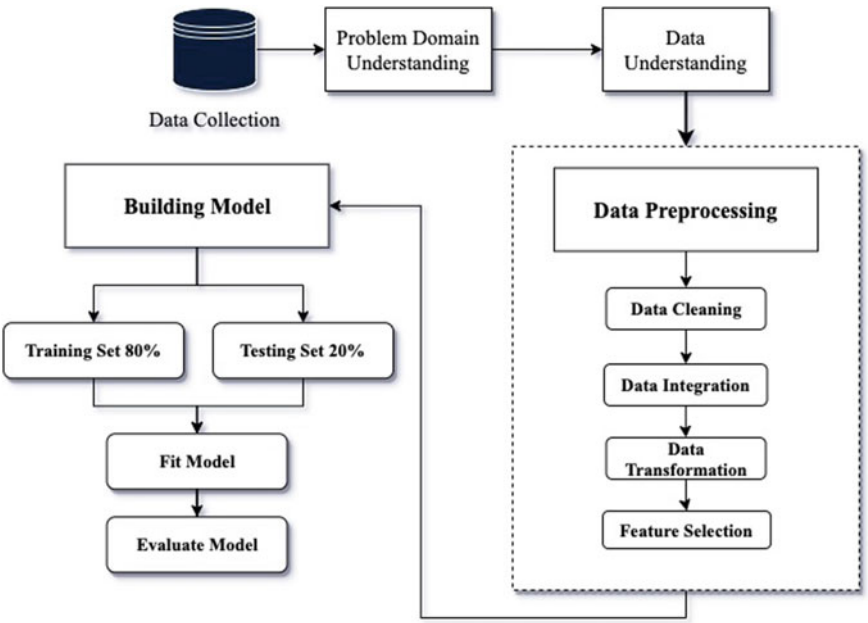


Fig. 1 Flow of building the proposed model

4.2 Understanding the Problem Domain

To understand the problem domain, despite studying many articles, several meetings have been initiated with the domain experts. Moreover, observations have been conducted with local health centers and many public hospitals to understand the current issues in the child's malnutrition screening process.

4.3 Understanding the Data

The collected dataset has belonged to the records of screened children from 2019 to the middle of 2020. The total collected data was 1,238 records that contained the nutrition-related features for children between the age of 1–59 months. Table 3 shows the description of all attributes of the target dataset. The dataset contained 13 attributes; the last attribute was considered as a label variable and consisted of 4 classes. The distribution of dependent variables was skewed or was not close to equal as depicted in Table 4, the majority class belonged to the (No) class.

4.4 Data Preprocessing

Since the data was collected from different sources, the data was combined for analysis. Fortunately, all the data had the same naming conventions for variables. Thus, the inconsistency was not seen to affect the reliability of the data. Further data

Table 3 Dataset features with short descriptions

No.	Feature	Type	Description
1	Entry_date	Date	The child's screening date
2	Age	Numeric	The age of the child in month
3	MUAC	Numeric	Mid-upper arm circumference in (mm)
4	Weight	Numeric	The weight of the child measured in (Kg)
5	Height	Numeric	The height of the child measured in (cm)
6	WHZ	Numeric	The weight-for-height z-score of the child
7	Cured	Binary	Whether the child is cured at the end of treatment
8	Death	Binary	The child is dead or not during the IPD treatment
9	Refer_date	Date	The date when the child is transferred out
10	Stay_length	Numeric	Total days of stay in hospital
11	Weight_gain	Numeric	Weight gain during the stay
12	Avg_weight_gain	Numeric	Average weight gain during the stay
13	Oedema	Nominal	Contained the oedema grades (0, g1, g2, or g3)

Table 4 Distribution of nutritional oedema classes in the dataset

No.	Class	Distribution Percentage (%)
1	No	59.17
2	Grade1	21.05
3	Grade2	14.01
4	Grade3	5.77
	Total	100

cleaning was carried out. Missing values were removed and some of them were replaced with the median. Moreover, the data was checked for extreme values and outliers, but the values were not found. In order to fit data into a suitable form for the model, the data was transformed to scale the data within a small range.

4.5 Building the Predictive Model

Once the data was preprocessed, three praiseworthy machine learning classifiers namely, Random Forest (RF), J48, and Naïve Bayes (NB) were applied and evaluated using the WEKA machine learning tool. The "InfoGainAttributeEval" and "ChiSquareAttributeEval" evaluation methods were used to evaluate the worth of each attribute for the model using the search ranker style. The dataset was split into 80% training set and the remainder for the testing set. For building the model, 6 out of 13 attributes namely, age, MUAC, weight, height, WHZ, and oedema were selected as the best features with a higher rank. The accuracy of classifiers was evaluated through the confusion matrix.

As often found in the real-world examples, the dataset was considerably imbalanced. Imbalanced data affected the accuracy of the model. Thus, more process was needed to get better classification categories in the data. For this purpose, Synthetic Minority Oversampling Technique (SMOTE) was applied to the training set of data to get almost balanced classification categories. SMOTE method over-samples the minority class by creating synthetic samples rather than replacement [20]. In the SMOTE operation, the resampling size was adjustable, in our experiment we started from 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, and 900% until we got the better resampled data to fit the model.

5 Results

This study aimed to identify the edematous malnutrition classes in under-five children. The proposed model predicts whether a child has grade1, grade2, grade3, or normal nutritional oedema. During experimentation, SMOTE (900%) with 5 nearest neighbors was applied. Three algorithms namely, RF, J48, and NB were

Table 5 RF, J48, and NB classifiers accuracy before and after SMOTE

No.	SMOTE	Accuracy (%)		
		RF	J48	NB
1	Before	69.92	69.07	68.64
2	After	**97.14**	94.51	54.17

used, and the results of each technique were compared. The experimental results of the three classifiers are presented in Table 5. The satisfying results were obtained up to 900% SMOTE, further resampling affected the accuracy. From Table 5 it was found that the RF, J48, and NB classifiers obtained 69.92%, 69.07%, and 68.64% accuracy, respectively before the SMOTE operation was applied.

When the SMOTE operation was applied, the RF, J48, and NB achieved 97.14%, 94.51%, and 54.17% accuracy, respectively. The experimental results provide evidence that SMOTE expectedly enhanced accuracy. Generally, the RF classifier delivered significantly better results after SMOTE (900%). Nevertheless, the NB classifier performed poorly. However, the result of the J48 classifier was close to the RF. A distinct improvement during the SMOTE operation was seen in the RF and J48. But the NB accuracy was not improved with SMOTE operation. Figures 2, 3, and 4 depict the variation of accuracy during the SMOTE operation ranging from 100% to 900% for each classifier.

6 Discussion

This study attempted to predict the edematous malnutrition of pre-school age children through non-clinical or anthropometric parameters in the context of Afghanistan. Oedema that causes SAM is one of the dangerous forms of malnutrition [2, 10]. The experimental data has consisted of four classes: oedema grade1, grade2, grade3, and normal. The design of the experiment was aimed to evaluate the effects of decision tree classifiers, when building the predictive tree models for multiclass classification problems using child's malnutrition data and to compare the performance of the RF, J48, and NB techniques to determine which one is performing well for predicting edematous malnutrition in under-five children.

In Table 5 in the results section, the accuracy for RF, J8, and NB classifiers before SMOTE was not satisfying due to the imbalanced data. Because feeding imbalanced data to the classifier makes it biased to the majority class and it is unable to learn about the minority class due to the insufficient data [20]. However, the RF and J48 after applying SMOTE yielded the most striking results. As depicted in Figs. 2, 3, and 4 in the results section, the SMOTE operation was applied on the training data started from 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, until 900%. In each step, SMOTE expectedly enhanced accuracy for RF and J48 classifiers. However, in Fig. 4 in the results section, SMOTE affected the performance of the NB classifier. Generally, from the results, it was found that

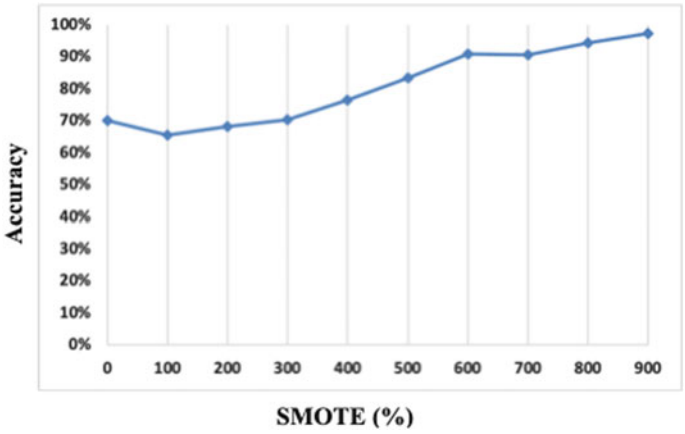


Fig. 2 RF classifier's variation of accuracy during SMOTE

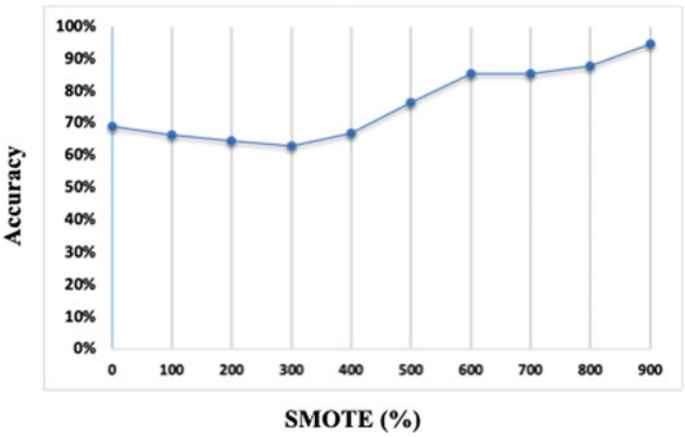


Fig. 3 J48 classifier's variation of accuracy during SMOTE

the SMOTE was a powerful approach for resampling our model's training data, eliminated the bias, and improved the accuracy of the model.

The incorporation of anthropometric data of a child to predict edematous malnutrition is a significant contribution of this research because oedema is a clinical indicator of malnutrition. The prediction of edematous malnutrition using machine learning has not been investigated by other studies [12–19]. The findings of this work indicate a well-built correlation between the anthropometric parameters of the child and clinical symptoms to identify SAM in children under five years of age. A basic shortcoming in this study is that there was limited dataset available for training and testing the model.

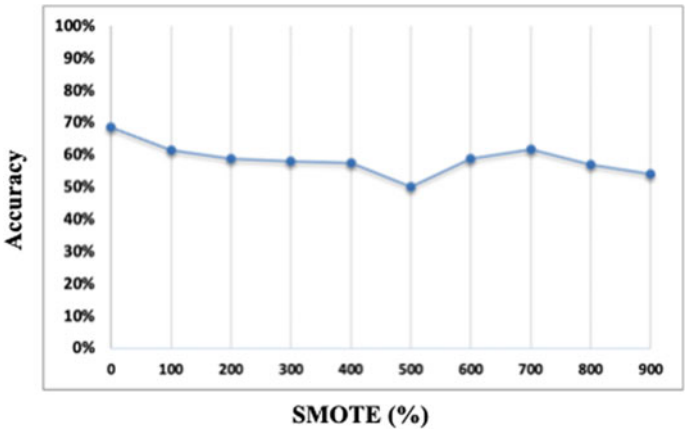


Fig. 4 NB classifier's variation of accuracy during SMOTE

7 Conclusion

Presently, in the healthcare research area, Artificial Intelligence and machine learning approaches are known as a vital tool for diagnosing various diseases. However, the potentials of machine learning have not been used to classify the edematous malnutrition in U5 children in Afghanistan. In this study, a model was developed to predict the edematous malnutrition status of pre-school age children. Three classifiers namely, RF, J48, and NB were used, and their results were compared. RF technique performed better than the other two classifiers. In future work, we will deploy the proposed model into a real mobile application to identify edematous malnutrition in children between the age of 1–59 months in Afghanistan.

References

1. Malnutrition in Children, https://data.unicef.org/topic/nutrition/malnutrition/. Last accessed 2020/12/2
2. Victora CG, Adair L, Fall C, Hallal PC, Martorell R, Richter L, Sachdev HS (2008) Maternal and child undernutrition: consequences for adult health and human capital. Lancet 371 (9609):340–357
3. Ministry of Public Health, Afghanistan (2018) Integrated management of acute malnutrition national guidelines. Public Nutrition Directorate, Kabul, Afghanistan
4. WHO Fact Sheets-Malnutrition. https://www.who.int/news-room/fact-sheets/detail/malnutrition. Last accessed 2020/12/3
5. UNICEF Afghanistan Nutrition. https://www.unicef.org/afghanistan/nutrition. Last Accessed 2020/12/21
6. Ministry of Public Health—Afghanistan (2004) Afghanistan National Nutrition Survey

7. Ministry of Public Health—Afghanistan (2013) Afghanistan National Nutrition Survey
8. Ministry of Public Health (MoPH), ICF, Central Statistics Organization (2017) Afghanistan Demographic and Health Survey, Kabul
9. Mother and Child Nutrition. https://motherchildnutrition.org/early-malnutrition-detection. Last accessed 2020/11/25
10. Everything you need to know about edema. https://www.medicalnewstoday.com/articles/159111#causes. Last accessed 2020/12/17
11. Momand Z, Mongkolnam P, Kositpanthavong P, Chan JH (2020) Data mining based prediction of malnutrition in Afghan children. In: 2020 12th international conference on knowledge and smart technology (KST). IEEE, pp 12–17
12. Khare S, Kavyashree S, Gupta D, Jyotishi A (2017) Investigation of nutritional status of children based on machine learning techniques using Indian demographic and health survey data. In: Procedia computer science, pp 338–349
13. Das S, Hossain MZ, Islam MA (2008) Predictors of child chronic malnutrition in Bangladesh. Proc Pak Acad Sci 45(3):137–155
14. Amato F, López A, Peña-Méndez EM, Vaňhara P, Hampl A, Havel J (2013) Artificial neural networks in medical diagnosis. J Appl Biomed 2(11):47–58
15. Nel, S., Feucht, U.D., Nel, A., Becker, P.J., Wenhold, F.A.M.: Ability of an Artificial Neural Network (ANN) to Predict Paediatric Severe Acute Malnutrition (SAM) from Routine Growth Curves. Available at SSRN 3728572
16. Shahriar M, Iqubal MS, Mitra S, Das AK (2019) A deep learning approach to predict malnutrition status of 0-59 month's older children in Bangladesh. In: 2019 IEEE international conference on industry 4.0, artificial intelligence, and communications technology (IAICT). IEEE, pp 145–149
17. Talukder A, Ahammed B (2020) Machine learning algorithms for predicting malnutrition among under-five children in Bangladesh. Nutrition 78:110861
18. Ariyadasa SN, Munasinghe LK, Senanayake SH, Fernando MGNAS (2013) Knowledge extraction to mitigate child malnutrition in developing countries (Sri Lankan Context). In: 2013 4th international conference on intelligent systems, modelling and simulation. IEEE, pp 321–326
19. Markos Z (2014) Predicting under nutrition status of under-five children using data mining techniques: the case of 2011 Ethiopian demographic and health survey. J Health Med Inform **5**(2)
20. Chawla NV et al (2014) Synthetic minority over-sampling technique. J Artif Intell Res 16:321–357

Discrete Particle Swarm Optimization for User Grouping in 5G NOMA Networks

Hadhrami Ab. Ghani, Farah Najwa Roslim, Eissa Mohammed Mohsen Al-Shari, Muhammad Akmal Remli, Nurul Izrin Md Saleh, Nooraini Yusoff, Hasyiya Karimah Adli, and Khairul Nizar Syazwan Wan Salihin Wong

Abstract The increasing demand for wireless network connections requires efficient network resource allocation. The non-orthogonal multiple access (NOMA) technology offers users sharing the same radio bandwidth to increase the bandwidth efficiency. However, the increase in the number of users demanding for the radio bandwidth and network connections will increase the required computational load for grouping the users to share the radio resources. This paper studies a heuristic method for grouping the users based on the discrete particle swarm optimization. The throughput, the average square error and the fitness function values obtained by the proposed method and the existing schemes are measured and observed. It has been demonstrated that the proposed scheme based on discrete particle swarm optimization has produced the throughput close to the upper limit. The convergence of the proposed method is mainly less than 10 iterations at different numbers of resource blocks.

Keywords Particle swarm · Discrete · User grouping

1 Introduction

High speed cellular networks in the future, which include 5G and beyond, are anticipated to offer higher network resources to an increasingly larger number of users. In providing services to more users, 5G networks have been proposed to be implemented with certain methods to ensure that the users are equally or fairly allocated with the available network resources. One of the most notable features in

H. Ab. Ghani (✉) · F. N. Roslim · E. M. M. Al-Shari · N. I. M. Saleh · N. Yusoff
Department of Data Science, Universiti Malaysia Kelantan, Kota Bharu, Malaysia
e-mail: hadhrami.ag@umk.edu.my

M. A. Remli · H. K. Adli · K. N. S. W. S. Wong
The Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, Kota Bharu, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_21

5G for utilizing and allocating the available network resources such as the bandwidth is the non-orthogonal multiple access system (NOMA) [1–5]. Instead of allocating dedicated bandwidth to the users, as what typically applied in the orthogonal multiple access system (OMA), a NOMA system allows the users to share the bandwidth [6]. Not only does it efficiently utilize and allocate the bandwidth [7, 8], but this mechanism also improves the effective sum capacity which can be achieved by the sharing users.

As the number of users grows bigger, the computational load required to perform the bandwidth allocation in NOMA systems tend to rise. Choosing the best users to be clustered in the same group which is allocated with the same radio bandwidth or frequency carriers requires a number of computations. This number is certainly dependent on the total number of users, which will be likely increasing. Therefore, it is important to devise computationally efficient methods for grouping the users and allocating them with the shared bandwidth. This is certainly timely and essential in improving the NOMA systems, which have been demonstrated to be viable in supporting the Ultra Reliable and Low Latency Communication (URLLC) and Massive Machine Type Communication (mMTC) [9] by permitting more users to share the resource blocks and network connections [10–12].

The computational load incurred in NOMA is also contributed by the implementation of successive interference cancellation (SIC) operation [13–17]. After detecting one of the users who share the same radio bandwidth, the SIC operation will be carried out to remove the signal of the users which has been detected to increase the probability of detecting the next user who shares the same bandwidth. Consequently, the effective sum capacity achievable by both users will be improved. Although the sum capacity improvement is achieved by using the SIC operation, the total required computational complexity will further increase as the number of cellular network users is on the rising trend [18–20]. Therefore, the need for a user grouping and bandwidth allocation mechanism with a relatively lower computational complexity is higher. The users that should be fairly allocated [21] with the same bandwidth should be chosen so as to increase the sum capacity using a computationally lower approach.

Due to complexity, most of the user grouping methods proposed in NOMA only consider user grouping, which allocates the shared bandwidth to a pair of selected users [22]. A user grouping method known as the power fixed fairness allocation (PFFP) is proposed in [22] to reduce the computational complexity and time. When compared with the exhaustive search (ES) scheme, which considers all possible pairs of users before determining the final pairs, this approach performs better in the complexity reduction. However, the PFFP approach has only considered a limited number of cellular network users. When the number of users increases, the computational complexity tends to rise, requiring computationally lower user grouping methods such as the heuristic methods which include particle swarm [23–25], ant-colony and drosophila optimization algorithms [26–31].

Although particle swarm optimization (PSO) has been implemented in 5G NOMA, most of the problems addressed are related to power domain NOMA and the corresponding power allocation. None of the proposed PSO methods addresses

the issue of grouping the users. Furthermore, the nature of PSO algorithm is not suitable for user grouping problems as PSO is naturally employed for solving continuous desired parameters such as the power allocation. In this paper, an improved PSO is designed for user grouping in NOMA for finding the best pairs of users to be allocated with the available bandwidth in the forms of resource blocks and frequency carriers. This discrete particle swarm optimization (DPSO) approach is designed with the SIC scheme to run a 5G NOMA user grouping model. Comparisons with the existing methods such as the PFFP and the ant-colony optimization approach will also be carried out.

2 Discrete Particle Swarm Optimization for User Grouping

A cellular network with N_{sites} cells and N_S sectors per cell is studied where NOMA is assumed to be implemented in each cell. There are N_{users} users in a cell and each of these users will be paired using the user grouping schemes that will be further described in this paper. Each of the pairs will be allocated with one resource block from a total of N_{r_b} resource blocks.

In the downlink direction, the total power allocated to a pair of two user signals is P_t and each user u will have an average transmitted power of $P_u = \alpha_u P_t$, with α_u represented as the allocated power ratio to user u from P_t, where $0 < \alpha_u \leq 1$ and $u \in \{1, 2\}$. The antenna gain and the fast fading function of user u at cell i and sector j are $G_{antenna}(j, i, u)$ and f_{j,i,u,r_b} respectively, where $j \in \{1, \cdots, N_S\}$, $i \in \{1, \cdots, N_{sites}\}$and $r_b \in \{1, \cdots, N_{r_b}\}$. The fading shadow and the path gain of user u at cell i are represented as $c_{u,i}$ and $G_{path}(i, u)$ respectively. Hence, the average received power of a user signal u at cell i and sector j is given as [22].

$$P_{j,i,u,r_b} = P_u G_{antenna}(j, i, u) f_{j,i,u,r_b} c_{u,i} G_{path}(i, u) \tag{2.1}$$

NOMA is implemented by first measuring the signal to interference plus noise ratio (SINR) of all users, assuming that the users are individually and independently allocated with resource blocks. In other words, this is an OMA setting is first and temporarily assumed without any bandwidth sharing as this individual SINR is required to group the users. When choosing the pairs, the first user must possess a higher SINR than the second user so that the first user can first be detected before SIC is performed to detect the second user. At the first cell and the first sector, the individual SINR $\gamma_{u,r_b}^{1,1}$ of user u is written as [22]:

$$\gamma_{u,r_b}^{1,1} = \frac{P_{1,1,u,r_b}}{\sum_{i=1}^{N_sites} \sum_{j=2}^{3} P_{i,j,u,r_b}}, \tag{2.2}$$

where the noise is assumed negligible in this interference-limited scenario.

By comparing between all SINR values, the first and the second user, which are going to be allocated with the same resource blocks, are selected by comparing between the SINR values. Therefore, the first and the second user are selected such that $\gamma_{u_1,r_b}^{1,1} > \gamma_{u_2,r_b}^{1,1}$, where both users are assumed to be allocated with the same transmit power for the purpose of determining the temporary OMA SINR and user selection. As the sharing users are now selected, the NOMA SINR value for the first user u_1 can be determined as follows:

$$\gamma_{u_1,r_b}^{1,1}(u_1) = \frac{P_{1,1,u_1,r_b}}{\sum_{i=1}^{N_{sites}} \sum_{j=2}^{3} P_{i,j,u_1,r_b}}. \tag{2.3}$$

The NOMA SINR $\gamma_{u_1,r_b}^{1,1}(u_1)$ for the first user u_1 appears to be very similar to that of OMA SINR calculated before because there is no SIC operation carried out on the first user during the detection process. However, the transmit power allocation for the first user is $P_{u_1} = \alpha_{u_1} P_t$ where $0 < \alpha_{u_1} \leq 1$. In other words, the total transmit power P_t is shared between the first and the second user, whose NOMA SINR $\gamma_{u_1,r_b}^{1,1}(u_2)$ is given as

$$\gamma_{u_1,r_b}^{1,1}(u_2) = \frac{P_{1,1,u_2,r_b}}{\sum_{i=1}^{N_sites} \sum_{j=2}^{3} P_{i,j,u_2,r_b} - P_{1,1,u_1,r_b}}. \tag{2.4}$$

Based on the SINR functions given in (2.3) and (2.4), the total mean throughput achieved by both users who share the same resource block can be written as

$$f(u_1, \ldots, u_{N_{ug}}) = 2W \sum_{n=1}^{N_{ug}} \log_2 \gamma_{u_n,r_b}^{s,c}(u_n). \tag{2.5}$$

2.1 Discrete Particle Swarm Optimization for User Grouping

The proposed algorithm based on discrete particle swarm optimization approach to group the users in 5G NOMA systems. As seen from Algorithm 1, the DPSO algorithm begins by populating S samples required before calculating the distances and velocity variables. Using an initialized set of distances $x_t(i,j)$ for $\forall i = 1, \ldots, S$ and $j = 1, \ldots, M$, the algorithm begins the for loop in line 3 to run the iteration up to N times.

At each iteration t, the distance will be recalculated and formulated in discrete form, as given in line 5, Eq. (2.7). The current best value will also be updated and the distance which corresponds to the current best value is updated in line 6, Eq. (2.8). If the current best value is larger than the global best value, the global best value will be updated accordingly, as shown in line 7 and 8.

From line 10 to 14, the velocity and the distance will be updated based on the general particle swarm optimization formula. The algorithm repeats itself until the maximum iteration has been reached. As explained before, the discretization of the distance happens in line 6. This signifies the main difference proposed in this paper than the other approaches implemented for the general particle swam optimization problems.

The discretization process will sort the values of the currently calculated distances in an ascending order. Once the distances are ordered, the indices of the ordered distances will be read and output as the discrete version of the distances. In literature there have been a number of approaches proposed to produce a discrete particle swarm optimization formula.

Some of the formula proposed in literature apply the exponential operation. Although this approach allows convergence of the learning process, the incurred computational load is relatively high due to the need to perform the exponential operation. For this reason, a more computationally lower approach is proposed in this paper to determine the discrete distance, hence completing the proposed DPSO algorithm.

Algorithm 1: Discrete Particle Swarm Optimization for NOMA User Grouping

Require: Maximum number of iterations, M; acceleration factor, $c_1,\ c_2$;inertia weight, w; population size, S; number of particles, M.

1: Generate S population samples.

2: Generate initialized distances $x_t(i,j)$ for $\forall\ i = 1,\cdots,S$ and $j = 1,\cdots,M$

3: for $t\ =\ 1{:}N$

4: **for** $i\ =\ 1{:}S$

5: $x_{d,t}(i,j) = \underset{j}{\text{sortindex}}\, x_t(i,j)$ (2.7)

6: $p_b\ =\ \arg\max_{x_t(i,j)} f\left(x_{d,t}(i,j)\right)$ (2.8)

7: **if** $f(p_b) > f(g_b)$ (2.9)

8: $g_b\ =\ p_b$ (2.10)

9: **end**

10: **for** $j\ =\ 1{:}M$

11: $v_t(i,j) =\ wv_{t-1}(i,j) + c_1 r_{1,t}\left(p_b - x_{t-1}(i,j)\right)$

12: $+c_2 r_{2,t}\left(g_b - x_{t-1}(i,j)\right)$ (2.11)

13: $x_t(i,j) = x_{t-1}(i,j) + v_t(i,j)$ (2.12)

14: **end**

15: **end**

16: **end**

17: **return** fitness function value, $f(g_b)$, user indices in each group

In the next section, the proposed DPSO algorithm will be implemented on a 5G NOMA model to group the users.

3 Results and Discussions

In order to test and validate the proposed scheme, a NOMA model is considered in a 5G cellular network. The total number of cells is 19 and the number sectors per cell is three. The gain and loss parameters and data used in this paper are the established data, as presented in [22], which follow the 3GPP standard. The output parameters of the tests and simulations run on the proposed DPSO scheme and the existing methods include the mean throughput, the number of iterations and the average mean square error per user, which has been given and described in the previous section. As for the DPSO parameter settings, the inertia weight, w is chosen between 0.9 and 1.2, the population size, S is varied between 8 and 50, the maximum iteration, N is between 3 to 100 and the acceleration parameters, c_1, c_2 are set to 2 [27].

The first test is run to measure the fitness function values achieved at each iteration for a range of N_{rb} values, from 1 to 5. The resource block number N_{rb} determines the number of users which can be paired. Therefore, 5 resource blocks can be allocated up to 10 users since each resource block will be shared by two users. Figure 1 shows the resulted fitness function values when the proposed DPSO method is run at iteration for different N_{rb} values ranging from 1 to 5. It general, the quickest convergence is achieved when $N_{rb} = 1$ and the slowest convergence occurs when $N_{rb} = 5$. The maximum fitness value is achieved at a value more than 0.9 when $N_{rb} = 5$ and the minimum fitness value is achieved at a value of about 0.2 when $N_{rb} = 1$. It can be also further observed that the maximum number of iterations before the graph curve is converged is 8, which is for the case of $N_{rb} = 5$, and the minimum number of iterations achieved before the graph is converged is 3, which is for the case of $N_{rb} = 1$.

The results obtained and shown in Fig. 1 shows that the convergence of the proposed DPSO is considerably quick. This is due to the discrete nature of the proposed DPSO algorithm used to group the users, leading to a faster convergence.

Apart from measuring the fitness function values, the mean throughput achieved is also measured. The proposed DPSO scheme has been run along with PFFP and ES schemes to record the mean throughput achieved in Mbps. As can be seen in Fig. 2, the proposed DPSO scheme has achieved the mean throughput close to the theoretical upper limit set by the ES scheme, which considers all options to perform the pairing of the users for allocating the resource blocks. The lowest mean throughput is recorded for PFFP method, although it has a relatively low computational complexity. It can be also observed that the proposed DPSO scheme has gained more than 4 Mbps as opposed to the PFFP scheme when $N_{rb} = 5$. This results further demonstrates the better performance of DPSO against the existing scheme, PFFP.

In Table 1, the maximum number of iterations recorded before the fitness function value converges is tabulated according to the corresponding number N_{rb} of resource blocks. It can be seen that the maximum number of iterations before convergence is 9, which occurs when $N_{rb} = 5$ which proves the effectiveness of the proposed scheme.

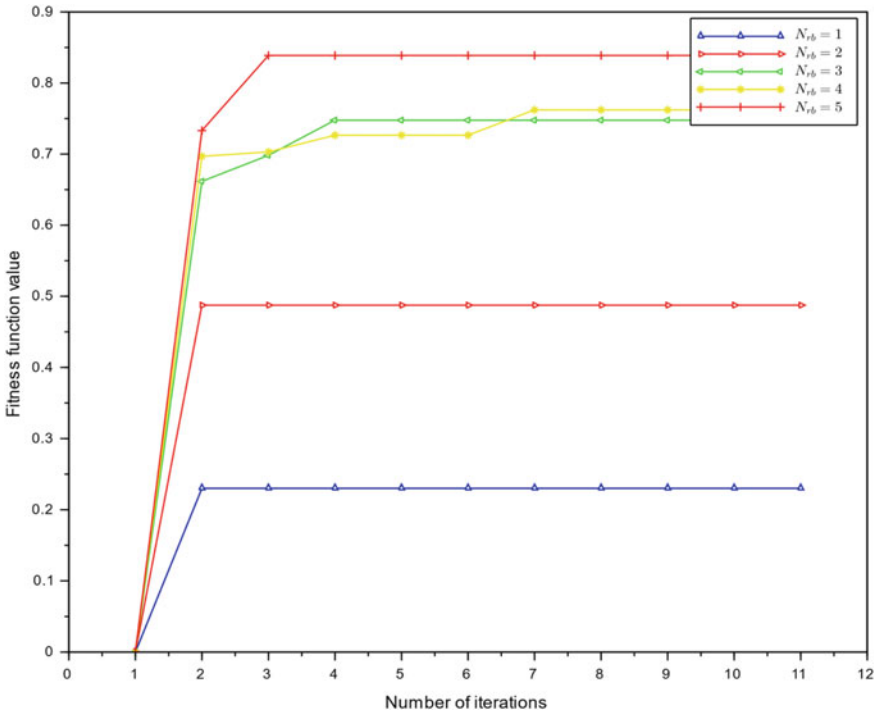


Fig. 1 Measurement of fitness function values at different N_{rb} values

In Fig. 3, the average mean square error (MSE) per user is measured as $MSE = 1/(1 + SINR)$ from all schemes under consideration when the number of resource blocks is between $N_{rb} = 1$ and when $N_{rb} = 5$. The mean square error is a good measurement of the performance of the SIC operation. As generally seen from the figure, the highest and the worst error is produced by the PFFP and the lowest error is achieved by the ES scheme. The proposed DPSO scheme follows closely the ES scheme in the error performance. This result further signifies the improved performance of the proposed DPSO as opposed to the existing PFFP scheme.

4 Conclusion and Future Work

In supporting the increasing number of users demanding the network connections from 5G networks and beyond, the proposed DPSO has been demonstrated to perform well in increasing the achievable mean throughput whiles reducing the average MSE per user. It has also been demonstrated that the required computational complexity by the proposed DPSO at each iteration is also low, hence

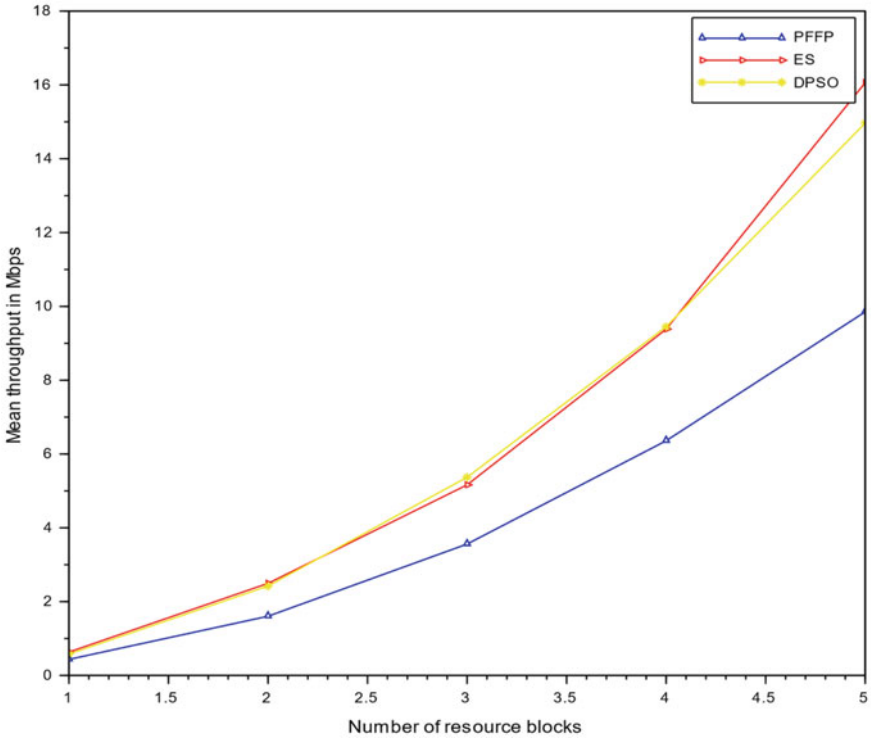


Fig. 2 Measurement of the mean throughput achieved by the proposed DPSO, PFFP and ES

Table 1 The maximum number of iterations before convergence for each N_{rb}

N_{rb}	Maximum number of iterations before convergence
1	2
2	2
3	4
4	7
5	9

potentially suitable to be implemented with more users. In the future, the proposed DPSO can be further enhanced and integrated with power domain NOMA systems which requires power allocation. A larger number of users can also be considered and compared against other heuristic approaches such as the ant-colony optimization method.

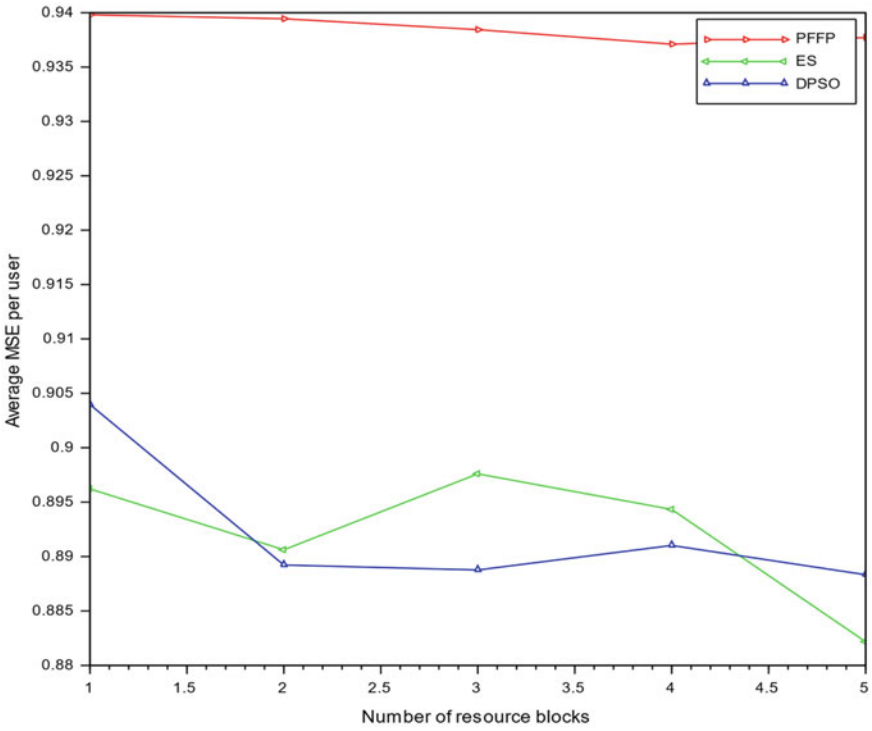


Fig. 3 Measurement of the average MSE per user

Acknowledgements This paper is prepared in accomplishment of a project at University Malaysia Kelantan (UMK) under a research fund, UMK-Fund Grant (R/FUND/A0100/01860A/001/2020/0082). The authors are hugely grateful to UMK and all parties who have helped to complete this paper.

References

1. Razavi R, Dianati M, Imran MA (2016) Non-orthogonal multiple access (NOMA) for future radio access. 5G mobile communications. Springer
2. Anwar A, Seet BC, Li XJ (2017) Interference modeling and outage analysis for 5G downlink NOMA. IEEE Veh Technol Conf. IEEE, Sydney
3. Ding Z, Dai H, Poor HV (2016) Relay selection for cooperative NOMA. IEEE Wirel Commun Lett 99:1
4. Kiani A, Ansari N (2018) Edge computing aware NOMA for 5G networks. IEEE Internet Things J 2018(5):1299–1306
5. Ali MS, Hossain E, Kim DI (2017) Non-orthogonal multiple access (NOMA) for downlink multiuser MIMO systems: user clustering, Beamforming, and power allocation. IEEE Access 99:1

6. Dai L et al (2018) A survey of non-orthogonal multiple access for 5G. IEEE Commun Surv Tutorials 20:2294–2323
7. Ding Z et al (2017) Application of non-orthogonal multiple access in LTE and 5G networks. IEEE Commun Mag 55:185–191
8. Islam SMR, Avazov N, Dobre OA, Kwak KS (2017) Power-domain non-orthogonal multiple access (NOMA) in 5G systems: potentials and challenges. IEEE Commun Surv Tutorials
9. ETSI et al (2018) ITU-R (2015) IMT vision—framework and overall objectives of the future development of IMT for 2020 and beyond, Recommendation, REC M. 2083-0, September 2015. Etsi Mec 0, 1–21
10. Venkatarman H, Trestian R (2017) 5G radio access networks
11. Dahlman E et al (2014) 5G radio access. Ericsson Rev (English Ed.) 91:42–47
12. Zaidi A et al (2018) Introduction: 5G radio access. 5G physical layer 1–19
13. McWade S, Flanagan MF, Zhang L, Farhang A (2020) Interference and rate analysis of multinumerology NOMA. In: IEEE international conference on communications
14. Mei W, Zhang R (2020) Cooperative NOMA for downlink asymmetric interference cancellation. IEEE Wirel Commun Lett 9:884–888
15. Herath P, Haghighat A, Canonne-Velasquez L (2020) A low-complexity interference cancellation approach for NOMA. In: IEEE vehicular technology conference 2020–May
16. Ding Z, Fan P, Poor HV (2017) Random Beamforming in millimeter-wave NOMA networks. IEEE Access
17. Mahady IA, Bedeer E, Ikki S, Yanikomeroglu H (2019) Sum-rate maximization of NOMA systems under imperfect successive interference cancellation. IEEE Commun Lett 23:474–477
18. Albashier MAM, Abdaziz A, Ghani HA (2018) Performance analysis of physical layer security over different error correcting codes in wireless sensor networks. In: International symposium on wireless personal multimedia communications, WPMC
19. Albashier MAM, Abdaziz A, Ghani HA (2017) Performance analysis of physical layer security over different t-error correcting codes. In: IEEE region 10 annual international conference, Proceedings/TENCON 2017–December, pp 875–878
20. Aziz AABD, Ghani HAB (2018) Energy efficiency in dynamic cluster selection for cooperative wireless sensor networks. In: 2018 IEEE region 10 symposium, Tensymp 2018, pp 155–159
21. Singh S, Buttar AS, Kaur D (2019) Survey on non orthogonal multiple access (NOMA)—a key technique for future radio network access. Int J Comput Sci Eng 7:794–799
22. He J, Tang Z, Che Z (2016) Fast and efficient user pairing and power allocation algorithm for non-orthogonal multiple access in cellular networks. Electron Lett 52:2065–2067
23. Strasser S, Goodman R, Sheppard J, Butcher S (2016) A new discrete particle swarm optimization algorithm. In: GECCO 2016—Proceedings of the 2016 genetic and evolutionary computation conference, pp 53–60
24. Hymes C, Klemmt C (2018) Discrete swarm logics. In: CAADRIA 2018—23rd international conference on computer-aided architectural design research in Asia: learning, prototyping and adapting, vol 1, pp 133–142
25. Dou J, Li J, Su C (2018) A discrete particle swarm optimisation for operation sequencing in CAPP. Int J Prod Res 56:3795–3814
26. Hani U, Samota KK (2019) Particle swarm optimization algorithm to improve access delay in 5G technology. Proceedings of the 2nd international conference on advanced computer control communication technology IAC3T 2018, pp 23–27
27. Masaracchia A, Da Costa DB, Duong TQ, Nguyen MN, Nguyen MT (2019) A PSO-based approach for user-pairing schemes in noma systems: theory and applications. IEEE Access
28. Mhudtongon N, Phongcharoenpanich C, Kawdungta S (2015) Modified fruit fly optimization algorithm for analysis of large antenna array. Int J Antennas Propag
29. Liyn LP et al (2020) Ant-colony and nature-inspired heuristic models for NOMA systems: s review. Telkomnika (Telecommunication Comput Electron Control) 18:1754–1761

30. Ghani HA, Aziz AA, Azizan A, Daud SM (2018) Adaptive interference mitigation with user grouping for fast transmission in cellular networks. Indones J Electr Eng Comput Sci
31. Ghani HA, Hamzah AA, Salem MA, Ahmed AA, Aziz AA, Azizan A (2020) Ant-colony optimization for 5G NOMA user grouping. IEEE international RF and microwave conference (RFM). IEEE, Sydney, pp 1–4

Ant-Colony Based User Grouping Modelling for Fast Cellular Networks

Ahmed Abdullah Ahmed Alhanik, Nur Asyiqin Hamzah, Hadhrami Ab. Ghani, Mohammed Ahmed Salem, Azlan Abdul Aziz, and Eissa Mohammed Mohsen Al-Shari

Abstract The need for efficient allocation of radio bandwidth to the users in the current and future cellular networks is high. As more users are requesting for the network connections, the radio bandwidth must be fairly and efficiently allocated. The non-orthogonal multiple access (NOMA), which has been introduced for 5G networks, is useful to allow bandwidth sharing among the users, rendering a more efficient allocation and utilization of the bandwidth. However, the computational load required to determine and select the users for each of the radio resources tends to increase as the total number of users rises. Therefore, user grouping that requires lower computational complexity such as the heuristic methods are useful to address this problem. An ant-colony optimization has been applied in this paper to perform user grouping in 5G NOMA. The mean throughput and mean square error have been measured when testing the proposed scheme. The result shows that the proposed scheme produces satisfactory throughput results, close to that of the theoretical upper limit. The mean square error is also close to the lower limit and better than the existing scheme.

Keywords Ant-colony optimization · User grouping · Non-orthogonal

1 Introduction

In today's new technology, wireless networking is a growing field. The ubiquity of wireless applications in different fields, such as financial transactions (banking), social networking, environment monitoring, device-to-device communications, etc., creates an open nature of wireless network broadcasting. Recently, there has been a

A. A. A. Alhanik · N. A. Hamzah · M. A. Salem · A. A. Aziz
Center for Engineering Computational Intelligence, Multimedia University, Malacca, Malaysia

H. Ab. Ghani (✉) · E. M. M. Al-Shari
Department of Data Science, Universiti Malaysia Kelantan, Kota Bharu, Malaysia
e-mail: hadhrami.ag@umk.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_22

substantial increase in user requirements for higher data rates, performance, better reliability and capacity growth. The key problems of wireless communication networks [1] are the management of these demands. The accessible design of wireless networking networks and the exponential increase in the number of users are causing these problems. The wireless network, for this purpose, needs more effective device scheduling to access the resource blocks. With the advancement of non-orthogonal multiple access (NOMA) wireless networking [2–5], one of the successful schemes to satisfy the rapidly rising user access criteria has been considered.

The orthogonal multiple access (OMA) techniques, such as orthogonal frequency division multiple access (OFDMA), time division multiple access (TDMA), and code division multiple access (CDMA) are effective at serving only one user in each orthogonal block. Non-orthogonal multi-access (NOMA) technology, however, is a critical support for wireless networks in the fifth generation (5G). The benefits of NOMA are greatly illuminated in fulfilling the heterogeneous demands for high throughput, high durability, enhance fairness, low latency and huge connectivity [6, 7]. The value of NOMA lies in a single resource block's capacity to support multiple users. In terms of spectrum, though NOMA is more effective, the network is susceptible to user intrusion from sharing.

More effective scheduling of users to access resource blocks is expected in the urban area, which is increasing rapidly in the number of users who can access the wireless network, causing higher computational complexity. Several mechanisms, such as round robin, are proposed for creating user groups. Since the systems involve a dramatically higher number of computations as more users are present, algorithmically lower approaches for classifying the groups of users have been proposed, such as heuristic models based on methods of artificial intelligence. Particle swarm optimization, dolphin echolocation optimization, drosophila optimization, optimization of ant-colony, firefly optimization, and genetic optimization are the heuristic models. In various fields, such as data analysis, computer vision, transportation and bioengineering these algorithms can solve challenges.

Ant Colony Optimization (ACO) is a population-based metaheuristic swarm intelligence based on the social actions of ant colonies that applies basic principles of distributed cooperation, self-organization, adaptation, and distribution found in ant populations to effectively solve real-life optimization problems. Rooting from the chemical concept of pheromones, which are deposited by the ants along the path they travel, ACO algorithm is useful for finding the best paths, groups or clusters based on a predefined objective function. It constitutes the big picture of the proposed ACO scheme in this paper before further elaborations are presented in other sections.

1.1 Research Gap

In wireless networking, the rising number of users is a significant and growing issue. By increasing the number of transmitters and receivers the capacity will get higher and as a result the interference and computational complexity will increase.

The interference was caused by the contact between antennas and several users positioned next to each other in the base station. As the number of users, who can access the wireless network, gets increase, a functional scheduling of the users to access the resource block is required. The scheduling operation for grouping these users should be carefully developed as the interference and computational complexity for grouping and scheduling will likely increase. The increase in interference in the grouping of a large number of cellular network users is worthy of consideration using the models of heuristic scheduling such as particle swarm optimization models (PSO) and ant-colony optimization (ACO). These strategies are developed for wireless and cellular networks to improve spectrum performance, enhance the throughput and decrease interference among users. The gap in this study is notably identified by proposing an integrated framework in mitigating interference mechanisms with ant-colony-based user grouping modelling to maximize mean throughput and minimize complexity.

1.2 Main Contributions

The main contributions of this study are as follows.

- A NOMA system for large number of users in 5G networks.
- An ACO-based user scheduling mechanism for assigning frequency whilst maintaining the optimal average throughput and decreased computational complexity.
- The effect on the mean throughput that caused by increasing the number of users in each resource block.

1.3 Organization

This paper also includes the following section. Section 2 points forth the models of heuristic. Section 3 displays the model of the system. Section 4 shows the effect of rising the number of users in each resource block. Section 5 presents the result and discussion of the paper. Finally, Sect. 6 present the conclusion of this paper.

2 Heuristic Models

The nature-influenced heuristic models studied here are the strategies that have been used in literature for seeking a solution to optimize pairs or groups of users on a wireless network.

2.1 Particle Swarm Optimization (PSO)

To solve the optimization issue, a useful technique that utilizes a probabilistic algorithm such like PSO can be used. It starts with a random solution to find an effective solution. This algorithm focuses solely on the ability of the groups based on a food process searching concept. PSO mimics swarm behaviors such as bird flocking and fish schooling. Bird can exchange information and find food with others [8]. PSO strategy has been used in a number of applications [9–11]. It is not difficult to execute and it is a combination of restricted parameters. PSO has been introduced for scheduling users via NOMA with its different sub-channels. PSO was often used to delegate power to each band to maximize the total throughput of the network as described in [12]. PSO has been implemented with ACO in [13] to determine whether this improves the efficiency which could be applied to evolutionary schemes. However, no work has been performed on other NOMA system models, such as ACO, though PSO does not work for big challenges.

2.2 Drosophila Optimization Model

Drosophila optimization is an ACO-like feature that is simple to use when working with practical problems. Drosophila is appropriate for multiple antenna systems and has been used in [14]. This model, along with NOMA, is going to play a very important role in the 5G, but drosophila is not suitable for scheduling users when grappling with certain claims.

2.3 Dolphin Echolocation Model

The Dolphin Echolocation (DE) principle is based on the action of dolphin while searching for food. DE can solve optimization problems by deciding the position of the target in the restriction area. This algorithm can also be used many times in different locations to include a variety of search spaces [15]. It is not necessary to use DE optimization for grouping users in pairs with small search spaces for multiple antennas.

2.4 Firefly Optimization Model

Firefly optimization model is inspired by the action of fireflies. This model is composed of a perception radius and a value of fluorescence used to quantify a single search and to determine the position output for a single search. Firefly goes to

the quest position by monitoring the light that another firefly creates to tick the target. Firefly has been used in the field of telecommunications, as stated in [16], but it has not been addressed in NOMA for users scheduling problems.

2.5 *Ant-Colony Optimization Model*

Ant-colony optimization (ACO) imitates rummaging behavior of ants for natural food sources [17]. Ants exchange information with others during their search expeditions using informal contact, such as "stigmergy" [18], and do so by producing traces of pheromones for other ants to detect food sources [19, 20]. ACO has been proposed to be integrated in various generations of cellular network technologies [21–23]. As the ACO scheme is also compliant with 5G, the purpose of this paper is to further extend and incorporate this 5G solution with a greater number of NOMA antennas and users. ACO is a method used to find an effective solution to optimization problems [24, 25].

3 System Model for User Scheduling and Grouping in NOMA

A 5G NOMA system is considered in the downlink direction where signals are transmitted from the base station to the receiver. There are N sites considered with three sectors for each site. The network sites are presumed to have a random and uniform number of users present. The mean power obtained per each resource block r is formulated as follows:

$$\mathcal{P}_{i,j,u,r} = \mathcal{P}_u G_{PL}(i,u) c_{i,u} G_A(i,j,u) f_{i,j,u,r} \tag{1}$$

The transmit power, $\boldsymbol{\mathcal{P}_u = \alpha_u \mathcal{P}_t}$, is evenly allocated per resource block (RB). In order to select the users to be grouped together, the individual signal-to-interference-and-noise ratio (SINR) is calculated for each user where no sharing of bandwidth is made, which is similar to the original Orthogonal Multiple-Access (OMA). The corresponding SINR, $\boldsymbol{\Upsilon_{u,r}^{1,1}}$ for Site 1 and Sector 1 of user $\boldsymbol{u}$ is written as follows:

$$\Upsilon_{u,r}^{1,1} = \frac{\mathcal{P}_{1,1,u,r}}{\sum_{i=1}^{N_s} \sum_{j=1}^{3} \mathcal{P}_{i,j,u,r} - \mathcal{P}_{1,1,u,r}} \tag{2}$$

Once the individual SINR (OMA) values have been determined, the grouping operation is started to choose and group the users. The needed signal for user $\boldsymbol{u}_1$ at the transmitting end is supplied with a transmitted power of $\boldsymbol{\alpha_{u_1} \mathcal{P}_t}$ and that of user $\boldsymbol{u}_2$ is $\boldsymbol{(1 - \alpha_{u_1}) \mathcal{P}_t}$, where $\boldsymbol{\mathcal{P}_t}$ is the total allocated power for each group, with $\boldsymbol{\alpha_{u_1} + \alpha_{u_2} = 1}$. After grouping the users, the corresponding SINR $\boldsymbol{\Upsilon_{u2,r}^{11}(u_1, u_2)}$ for

user $\boldsymbol{u_2}$ is calculated as:

$$\Upsilon_{u_2,r}^{1,1}(u_1,u_2)=\frac{\mathcal{P}_{1,1,u_2,r}}{\sum_{i=1}^{N_s}\sum_{j=1}^{3}\mathcal{P}_{i,j,u_2,r}-\mathcal{P}_{1,1,u_2,r}+\mathcal{P}_{1,1,u_2,r}} \tag{3}$$

where the expression of the **SINR** for user $\boldsymbol{u}_1$ is:

$$\Upsilon_{u_1,r}^{1,1}(u_1,u_2)=\frac{\mathcal{P}_{1,1,u_1,r}}{\sum_{i=1}^{N_s}\sum_{j=1}^{3}\mathcal{P}_{i,j,u_1,r}-\mathcal{P}_{1,1,u_1,r}} \tag{4}$$

Interference caused by user $\boldsymbol{u_2}$ is then removed using a SIC operation before detecting user $\boldsymbol{u_1}$, which has a better SINR without $\boldsymbol{u_2}$. Therefore, the objective of the grouping algorithm based on ACO is to optimize the throughput for all groups of users, which can be expressed as:

$$R_{t,u,g}=\sum_{i=1}^{N_{\mathrm{ug}}}\Upsilon_{u_i,r}^{c,s}\left(u_1,..,u_{N_{\mathrm{ug}}}\right) \tag{5}$$

4 ACO-Based User Scheduling and Grouping Algorithm

The suggested user grouping approach which is based on ACO is designed by segregating users into groups or pairs which are allocated with the available radio resources. In order to apply the ACO, $\boldsymbol{N_{ant}}$ ants are generated to find the best groups for each user, with $\boldsymbol{N_{user}}$ users are present and $\boldsymbol{N_u}$ users in a group, the number of groups is determined as:

$$N_{ug}=\frac{N_{user}}{N_u} \tag{6}$$

The ants which act as the agents finding the best routes, or essentially groups will leave behind the pheromone along the routes that have been gone through. The more the pheromone value is, which depends on the number of ants as well as the evaporation rate, the more likely that the routes, or the groups, will be chosen as the groups for the target users, based on the objective function value $R_{t,u,g}$. The two methods of ACO optimization algorithms, ACO and ACS, as presented in [24],

Table 1 Computational complexity of the existing schemes

Schemes	Computational complexity	
ACO	$N_{ant}\left(\sum_{k=1}^{N_{user}}k\right)$	(7)
ACA	$N_{ant}\left(\sum_{k=1}^{N_G-1}\binom{N_u+kN_u}{N_u}\right)$	(8)
SHS	$N_{iter}\left(\frac{N_{user}^2}{N_u^2}+N_{user}\left(\sum_{k=1}^{N_G-1}\binom{kN_u}{N_u-1}\right)\right)$	(9)

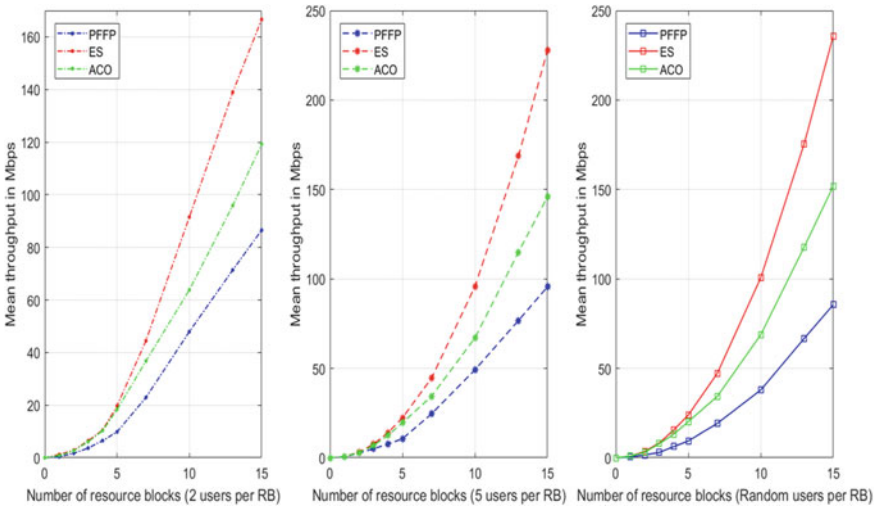


Fig. 1 The mean throughput Mbps of the Large antenna system

Table 2 The mean throughput achieved by applying the ACO technique at 15 RB

Number of users per RB	Throughput (Mbps)
2	119.2164
5	145.872164
Random (more than 5)	151.872164

have been applied and tested for grouping. Table 1 tabulates the computational complexity for the two ACO-based algorithms along with another grouping method, SHS, which is based on binary-swapping concept. It shows that the ACO method performs better than the other two methods with lower complexity.

5 Results and Discussions

The proposed ACO user grouping method is tested to determine the achievable throughput in a 5G NOMA system. For a comparison purpose, two other existing schemes have been run and tested, which are the ES and PFFP scheme [25]. The comparison is carried out to evaluate the performance of users grouping by grouping the users with one resource block allocated to each pair or group. Figure 1 the first result on the achievable throughput by the three schemes including the proposed ACO method is shown where the number of resource blocks N_{RB} is varied between 1 and 15.

As can be seen the throughput performance recorded in Fig. 1, the main performance rises steadily when the resource blocks are between 1 and 5. When the

N_RB resource block between 5 to 15 a more important change in the efficiency may be found. From Fig. 1, it can be observed that the performance of PFFP is lower as compared with ACO, especially when the number of resource blocks increases. The achievable performance of ACO throughput is very close to ES, which stands as the upper bound.

Moreover, the impact of increasing the users per resource block is investigated on the mean throughput. Table 2 illustrates the mean throughput achieved by applying the ACO technique at 15 RB with a different number of users.

Based on Table 1 and Fig. 1, we observe that by increasing the number of users per resource block the performance of the ant-colony optimization further improves. This observation is very significant as the number of resource blocks increases in inevitable due to the increasing demand from the users.

6 Conclusion

The ant-colony optimization scheme has been developed and implemented for grouping the users in 5G NOMA networks. The results show that the performance of the proposed scheme is close the theoretical exhaustive search, which considers all options to find the highest possible throughput. This is essential as the total number of users tend to increase and without a computationally efficient approach like this ant-colony optimization scheme, the required and incurred computational load may significantly increase and affect the whole performance of the network.

Acknowledgements This project is funded under a FRGS Project Grant (Grant no. FRGS/1/2019/TK04/MMU/02/4) run in Multimedia University Melaka.

References

1. Aziz AA, Ghani HA (2018) Energy efficiency in dynamic cluster selection for cooperative wireless sensor networks. In: 2018 IEEE region ten symposium (Tensymp), pp 155–159
2. Ahmed MA, Mahmmod KF, Azeez MM (2020) On the performance of non-orthogonal multiple access (NOMA) using FPGA. Int J Electr Comput Eng 10(2):2151–2163
3. Buskgaard E, Tatomirescu A, Del Barrio SC, Franek O, Pedersen GF (2014) Effect of antenna bandwidth and placement on the robustness to user interaction. In: 2014 Int Work Antenna Technol Small Antennas Nov EM Sruct Mater Appl iWAT 2014, pp 258–261
4. Mahady IA, Bedeer E, Ikki S, Yanikomeroglu H (2019) Sum-rate maximization of NOMA systems under imperfect successive interference cancellation. IEEE Commun Lett 23(3):474–477
5. Kuo Y, Structure AO (2020) Research and design of a high isolation 5G antenna for smart phone. In: 3rd international conference on information and computer technologies (ICICT), pp 507–510
6. Yu Y, Chen H, Li Y, Ding Z, Song L, Vucetic B (2018) Antenna selection for MIMO nonorthogonal multiple access systems. IEEE Trans Veh Technol 67(4):3158–3171

7. Hasan MK, Shahjalal M, Islam MM, Alam MM, Ahmed MF, Jang YM (2020) The role of deep learning in NOMA for 5G and beyond communications. 2020 Int Conf Artif Intell Inf Commun ICAIIC 2020:303–307
8. Anantathanavit M, Munlin MA (2013) Radius particle swarm optimization. In: 2013 Int Comput Sci Eng Conf ICSEC 2013, pp 126–130
9. Hani U, Samota KK (2019) Particle swarm optimization algorithm to improve access delay in 5G technology. Proc. 2018 2nd Int Conf Adv Comput Control Commun Technol. IAC3T 2018, pp 23–27
10. Guo YX, Li H (2018) A power allocation method based on particle swarm algorithm for NOMA downlink networks. J Phys Conf Ser 1087(2):5–12
11. Masaracchia A, Da Costa DB, Duong TQ, Nguyen MN, Nguyen MT (2019) A PSO-based approach for user-pairing schemes in NOMA systems: theory and applications. IEEE Access 7:90550–90564
12. Pliatsios D, Sarigiannidis P (2019) Power allocation in downlink non-orthogonal multiple access IoT-enabled systems: a particle swarm optimization approach. Proc 15th Annu Int Conf Distrib Comput Sens Syst DCOSS 2019, pp 416–422
13. Selvi V, Umarani DR (2010) Comparative analysis of ant colony and particle swarm optimization techniques. Int J Comput Appl 5(4):1–6
14. Mhudtongon, N., Phongcharoenpanich, C., Kawdungta, S.: Modified fruit fly optimization algorithm for analysis of large antenna array. Int J Antennas Propag
15. Poornima N, Santhi M, Seetharaman G (2019) Dolphin echolocation based generation of application definite Noc custom topology. Int J Recent Technol Eng 8(3):8247–8254
16. Mosavvar I, Ghaffari A (2019) Data aggregation in wireless sensor networks using firefly algorithm. Wirel Pers Commun 104(1):307–324
17. Liu, L., Dai, Y., Gao, J.: Ant colony optimization algorithm for continuous domains based on position distribution model of ant colony foraging. Sci World J
18. Pedemonte M, Nesmachnow S, Cancela H (2011) A survey on parallel ant colony optimization. Appl Soft Comput J 11(8):5181–5197
19. Dorigo, M., Maniezzo, V., Colorni, A.: Ant system: Optimization by a colony of cooperating agents. In: IEEE transaction on system man, cybernetics part B cybernetics, vol 26, no 1, pp 29–41
20. Ping G, Chunbo X, Yi C, Jing L, Yanqing L (2015) Adaptive ant colony optimization algorithm. Proc 2014 Int Conf Mechatronics Control ICMC 2014, no. Icmc, pp 95–98
21. Sefako T, Walingo T (2020) Biological resource allocation algorithms for heterogeneous uplink PD-SCMA NOMA networks. IEEE Access, pp 1–1
22. Liu Y, Tao M, Li B, Shen H (2010) Optimization framework and graph-based approach for relay-assisted bidirectional OFDMA cellular networks. IEEE Trans Wirel Commun 9 (11):3490–3500
23. Bin Hamzah MH, Bin Ab Ghani H (2014) Energy-limited ant colony optimization with cooperative transmission. I4CT 2014 1st Int Conf Comput Commun Control Technol Proc, no. 14ct, pp 161–165
24. Ghani HA, Hamzah MH, Syahali S, Aziz NHA (2016) Ant-colony algorithm with interference cancellation for cooperative transmission. IET Signal Process 10(6):603–610
25. Ghani HA, Azizan A, Aziz A, Daud SM (2018) Adaptive interference mitigation with user grouping for fast transmission in cellular networks. Int J Electr Eng Comput Sci 10(2):704–712

A Systematic Approach to Transform Machine Learning Students' Performance Prediction Model into Preventive Procedures

Ijaz Khan, Abdul Rahim Ahmad, Nafaa Jabeur, and Mohammed Najah Mahdi

Abstract The modern day educational institutes are craving novel procedures to utilize the students' data to amplify their prestige and improve the education quality. One of the major problems an instructor/institute experiences is the thorough monitoring of students' academic progress, in a course, and instigate preventive procedures to offer additional support to the students with unsatisfactory academic progress. Educational Data Mining tools, specifically Machine learning classifiers, appear supportive to develop prediction models which forecast students' final outcome in a course. This research evaluates the effectiveness of machine learning classifiers to monitor students' academic progress and informs the instructor about the students at the risk of producing unsatisfactory final result in a course. The dataset is pre-processed with Pearson correlation feature selection algorithm to discover the features which influence the students' academic performance. A set of machine learning models are developed and compared through accuracy, sensitivity, specificity, F-measure and Mathew Correlation Coefficient, to choose the finest model. J48 decision tree prevails other models by achieving accuracy and F-measure of nearly 0.90. Simple logistic appeared the least effective model while an altered version of k-nearest neighbor achieved highest sensitivity but remain ineffective due

I. Khan (✉)
College of Graduate Studies, Universiti Tenaga Nasional, Kajang, Malaysia
e-mail: ijaz@buc.edu.om

I. Khan
Information Technology Department, Buraimi University College, Al-Buraimi, Oman

A. R. Ahmad
College of Computing and Informatics, Universiti Tenaga Nasional, Kajang, Malaysia
e-mail: abdrahim@uniten.edu.my

N. Jabeur
Computer Science Department, German University of Technology, Muscat, Oman
e-mail: nafaa.jabeur@gutech.edu.om

M. N. Mahdi
Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Kajang, Malaysia
e-mail: najah.mahdi@uniten.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_23

to lower accuracy and F-Measure. The ideal model is further transformed into easily explicable format and then interpreted into a set of supportive measures to carefully monitor students' performance from the very start of the course and a set of preventive measures to offer additional attention to the struggling students.

Keywords Educational data mining · Machine learning · Decision tree · Student performance prediction

1 Introduction

The educational institutes are more apprehensive about their prestige as compare to the quality of education [1]. However, various government and accreditation agencies ensure the educational institutes sustain a high quality learning environment and the concrete procedures of accreditation has compelled the institutions to plan and implement novel procedures to preserve their standards. Therefore, the modern-day educational institutes are trying to uphold quality and prestige in the education society. To maintain the standards, the institutes are in quest of innovative practices to conserve their position.

The institutions put novel technologies into practice to facilitate them to congregate vast amount of data about the learners and the underlying learning environment [2]. Longitudinal studies reveal that the modern-day student owns an arbitrary amount of technologies, for instance personal computer, laptop, Smartphone, maybe tablet or an e-Reader. In combination with a mobile internet connection, it is reasonable to say that practically everyone has access to information and communication in real time even while on the move. Further, modern web technologies also allow to develop innovative information systems that are able to store huge amount of data [3].

The institutes require inventive procedures to utilize the collected data and enhance their potential to make more constructive decisions. Data Mining (DM), encompassing computational techniques, analyze the data and discover valuable hidden patterns [4]. Educational Data Mining (EDM), an emerging discipline of Data Mining, collect data about learner and their learning environment, and propose novel methods to identify useful patterns within this data which appear helpful in augmenting learner, learning environment and the related factors. A number of computer technologies offers convenience to renovate the complex material to make it easy to understand and remember [5]. Machine Learning is among the tools possessing the potential to support the education institutions in several states of affairs. The machine learning algorithms make use of the previous data and forecast the likelihood of an event with suitable precision [6].

One of the major problems an instructor faces is to determine the level of students at the start of the course. The instructor can devise essential preventive measures to deal with the students struggling in the course. This leads towards the need of novel measures to predict the struggling students and make it possible for

the instructor to initiate preventive measures. Novel prediction models are imperative to track students' academic performance, in a course, and identify the students with inadequate academic progress [6]. A number of prediction models, such as [7, 8] have been implemented to back up machine learning algorithms as useful tools to predict students' final outcome. Several prediction models, for example [9, 10], monitor students' performance in online learning. The models, for instance [11], identify the students who are more likely to drop the course at the beginning of the semester. Several authors design models [6, 12] specifically to track the performance of novice computer language programmers.

The prediction model is inadequate without utilizing it to devise precautionary measures. The existing models are productive solely in the local context. Learning is always local and the students within different educational environments may respond in a different way [13]. This research work applies machine learning algorithms, decision tree, Naïve Bayes, Simple Logistic and k-nearest neighbors, to develop prediction models which are able to recognize the students with poor academic performance at an early stage of the semester in a course. Once an appropriate model is selected, then it is transformed into easily understandable form so the instructor, even with little computer knowledge, can understand and interpret it into supportive procedures. The final product of the model comprises a set of preventive measures to support the inadequate students. The paper is organized as; Sect. 2 provides a literature review of educational data mining, machine learning algorithms and discusses several student performance prediction models. Section 3 provides the 3-step methodology used in this research and explains each phase in details and Sect. 4 concludes the paper.

2 Literature Review

Educational Data Mining (EDM) makes use of student data; demographic information, presage academic record, current academic behavior, and data obtained from students' management systems to explore the unique types of data that come from an educational context and understand students and the settings in which they are learning [14]. The discovery of underlying activities aids to pull out constructive information that can support the institution's decision making. Provided that proper practices are in-placed, the data can be utilized to identify the students who deviate from satisfactory academic performance.

Machine Learning algorithms possess the ability to enable computers to learn automatically and set the procedures to make future decisions [15]. Machine Learning prediction models predict the final outcome of the students based on several academic features. The main output of the model identifies the students likely to end with unsatisfactory outcome. These struggling students can be forwarded for supportive and counseling mechanisms. Numerous machine learning, mainly supervised algorithms are used to put into practice the notion of student performance prediction modeling.

Numerous models have been proposed under different educational contexts to address the student performance prediction. Maesya and Hendiyanti [16] developed model to forecast if the student will graduate on time or late than the standard graduation duration. Asif et al. [7] used machine learning classifiers to develop a model to predict the final result of students at an early stage of the study and thus facilitates the management to focus more over the bright students and support the weak students. Costa et al. [12] developed models to identify students having higher probability to fail the introductory programming courses. Orong et al. [17] used modified Genetic Algorithm to eliminate excessive features and applied decision tree algorithm to discover the weak students and facilitate the institution to design interference measure to raise the student attrition. Chen [18] built models with decision tree and linear regression with a set of features extorted from the institution's auto-grading system to assist the institution in recognizing the struggling students and assigning teaching hours automatically in a smart way. Saa [19] proposed a decision tree model to discover the essential features which influence students' academic performance. There are a number of models which are based on decision tree [20], lazy classifiers [21], Artificial Neural Networks [11, 22]. Naïve Bayes [23, 24] and Support Vector Machine [25, 26].

The literature confirms machine learning algorithms as productive tools for developing models to predict students' final outcome. This research aims to develop prediction model and then transform the model into a form which is easily understandable by the instructor. The transformation yields sets of precautionary measure for the students. This model proposes appropriate measures for prior and post model execution.

3 Research Methodology

Figure 1 illustrates the 3-phased research methodology used in this research. The first phase manages to clean the data and perform necessary pre-processing. The primary task is to describe and prepare the training dataset. Data Pre-Processing, enhances the quality of dataset so the machine learning algorithms can produce

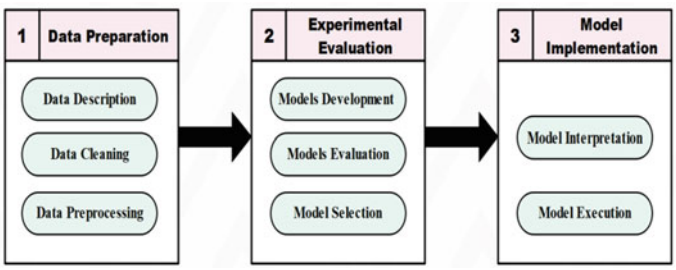


Fig. 1 The methodology used in this research

better results. The experimental evaluation phase applies several machine learning algorithms over the dataset and produces a prediction model for each algorithm. The model which better suits the addressed context is chosen for interpretation. The chosen model is transformed to an effortlessly understandable form in model implementation phase. The final step proposes precautionary measures with regards to the transformed model.

3.1 Data Preparation

Data Description. The data, in this research, is collected from the registration department of Al-Buraimi University College (BUC), Sultanate of Oman with consideration of entire ethical guidelines. The dataset comprises of the students enrolled in a course in the past 3 semesters.

Data Cleaning. Data cleaning organizes the dataset by dealing with missing value, data noise and resolving irregularity in the dataset [27]. We eliminate the irrelevant features, for instance student's college ID. Noisy data, in fact, downgrades the machine learning algorithm's performance and therefore these features need to be carefully dealt either by removing them from the dataset or using proper noise removal techniques.

Data Pre-Processing. High dimensionality and class imbalance are the major issues, related to training dataset, which can demean the performance of the prediction model [28]. This pre-processing task can be achieved by applying feature selection algorithms. We make use of Correlation Attribute Evaluator Filter with Ranker search to reduce the number of attributes. It measures the Pearson's correlation between each attribute and the prediction class and the unrelated features turn up with low correlation value. Table 1 provides the list of features along with their description and their correlation value with the prediction class.

In our dataset, the minority class has a very low number of instances; therefore, we apply Synthetic Minority Over-sampling Technique (SMOTE) to increase the number of instances in the minority class. The dataset, after applying SMOTE, consists of 223 instances with approximately even class distribution.

3.2 Experimental Evaluation

Model Development. The experiments are carried out in Waikato Environment for Knowledge Analysis (WEKA) [29]. Four state of the arts machine learning algorithms from distinct categories, J48, 15-NN, Naive Bayes and Simple Logistic, are applied over the training dataset to produce prediction models. We employed

Table 1 The list of student features with their description

Features	Values	Pearson correlation value
CGPA	Real	0.524
PreReq_Grades	Real	0.489
Grade-1_Cont	Real	0.445
Attendance	Real	0.343
Gen	Nominal (Male, Female)	0.158
Dorms	Nominal (Yes, No)	0.134
Session	Nominal (Morning, Evening)	0.072
Year	Nominal (From one to four)	0.068
Sponsorship	Nominal (Yes, No)	0.033
Major	Nominal (Translation, Literature)	0.023
Final_Grade	Prediction Class (Low and High)	

Table 2 Confusion matrixes of the prediction models

	J48		15-NN		Naïve Bayes		Simple logistic	
Classified as	High	Low	High	Low	High	Low	High	Low
High	103	12	89	26	95	20	91	24
Low	12	96	6	102	8	100	9	99

tenfold cross-validation [30]. We make use of Confusion Matrix to visualize the prediction models. Table 2 provides the confusion matrixes for the developed models.

Models Evaluation. Several evaluation metrics, with distinct features, can be computed to assess the performance of classifiers. The competency to accurately classify instances into their correct classes illustrates the domination of the model. We make use of accuracy, precision, sensitivity (recall), specificity, F-measure and Mathew Correlation Coefficient (MCC). MCC [31] evaluates the performance of the classifier in terms of how well it classifies the instances in correct classes.

$$\begin{aligned}
\text{Accuracy} &= (TP + TN)/(TP + FN + FP + TN) \\
\text{Sensitivity (Recall)} &= TP/(TP + FN) \\
\text{Specificity} &= \text{TN}/((\text{TN} + \text{FP})) \\
\text{F-Measure} &= 2*(\text{Precision}*\ \text{Recall}\ )/(\text{Precision} + \text{Recall}) \\
\text{Precision} &= \text{TP}/(\text{TP} + \text{FP}) \\
\text{MCC} &= \frac{\text{TP.TN} - \text{FP.FN}}{\sqrt{(\text{TN} + \text{FN}).(\text{TP} + \text{FP}).(\text{TN} + \text{FP}).(\text{TP} + \text{FN})}}
\end{aligned}$$

Figure 2 illustrates J48 achieving an accuracy of near to 0.90 followed by Naïve Bayes with 0.87. This shows that the prediction models are showing an excellent

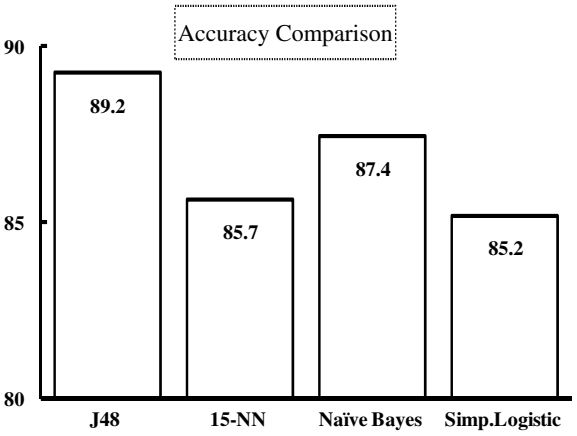


Fig. 2 Comparison of accuracy

understanding of the training dataset. Accuracy alone does not assure the dominance of the prediction model. Therefore, we extracted sensitivity and specificity of the models, in Fig. 3, to get a broader view of the models' behaviour. Figure 3 demonstrates J48 achieving highest sensitivity, while 15-NN achieves the highest specificity. Despite having highest specificity, 15-NN achieves the lowest sensitivity. Similarly, J48 achieves the lowest specificity.

F-measure provides an indication of overall performance of the model in terms of precision and recall values. Figure 4 compares the F-measure values of the models. It shows J48 is achieving the highest F-measure values followed by Naïve Bayes. It shows 15-NN, despite having highest specificity, achieves lowest F-measure value. Figure 5 illustrates J48 leaving behind the entire set of algorithms in MCC. This indicates that J48 has higher ability to accurately place the instances in their respective classes.

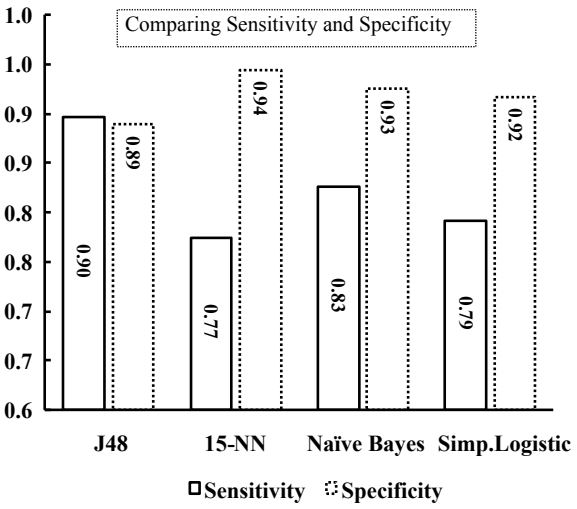


Fig. 3 Comparing precision and recall

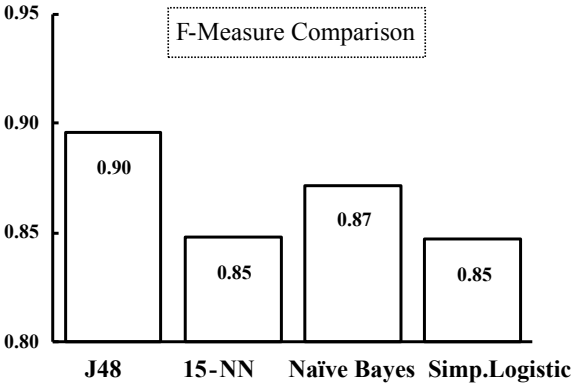


Fig. 4 Comparison of F-measure values

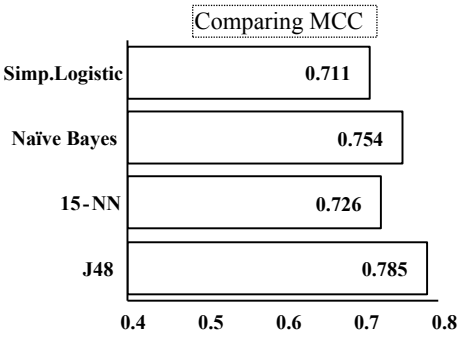


Fig. 5 Comparison of MCC

Model Selection. A higher sensitivity and specificity illustrates the models capability to identify majority and minority class instances respectively. This indicates, 15-NN appears handy if the main objective of the instructor is to segregate the students with low academic performance and J48 appears a good choice if the objective is to identify the student in the "high" class.

The comparative analysis shows J48 model leading in accuracy and F-measure values. Achieving high accuracy illustrates its capability to correctly classify True Positive (TP) and True Negative (TN) instances. Similarly, the high F-measure further favours J48 dominancy while 15-NN stays lower in both accuracy and F-measure comparison. MCC highlight the capability of J48 to classify the instances correctly within both classes which further corroborate J48 pre-eminence. Decision tree classifiers provide clear illustrations that are easily understandable even by ordinary users. Therefore, we propose to choose decision tree based model for further evaluation.

3.3 Model Implementation

Model Interpretation. The previous section (model selection) concludes decision tree (J48) as an appropriate model in the current context. The rules from the decision tree further clarify the model. Figure 6 depicts the decision tree as extracted from WEKA. The decision tree makes use of CGPA, midterm exam marks (Grae-1_Cont), attendance and grades in the prerequisite subject (PreReq_Grades). CGPA stays at the root of the tree and thus is the prime feature of students' performance. Figure 7 provides the rules extracted by J48 from the training dataset. It provides a better view of the decision tree for programming the model.

Model Execution. The goal is to transform the model in easy to understand form and design preventive actions for the students with inadequate performance. The output model reveals attendance as one of the key factor and thus the instructor should constantly emphasis over its significance from the start of the course/ semester. In order to lessen the consequences of absence, the instructor can revise the teaching methodology by adding time slot for quick revision of the previous class. These measures can reduce the number of inefficient students prior to the midterm exam.

The model is executed immediately after the midterm exam to obtain a list of students with their probable final outcome. The student in real danger of failing the course will be forwarded for further advisory consultations. The instructor may interview them and plan precautionary actions compliant with the adapted perspective of every student. Additionally, the instructor may arrange additional classes in an attempt to bring the at-risk students on right track. In additional classes, the instructor should revise the course contents and motivate the students.

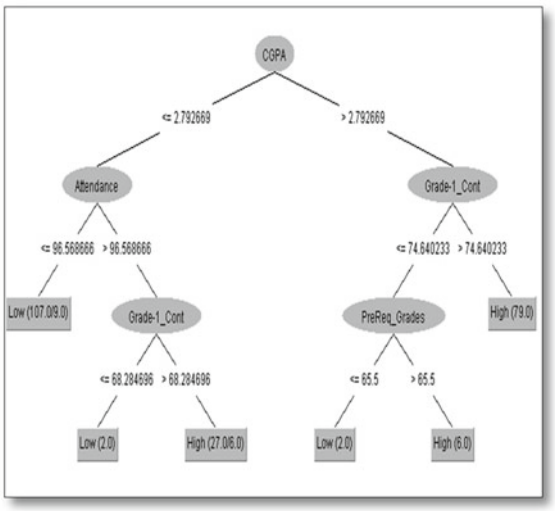


Fig. 6 Extracted decision tree (J48)

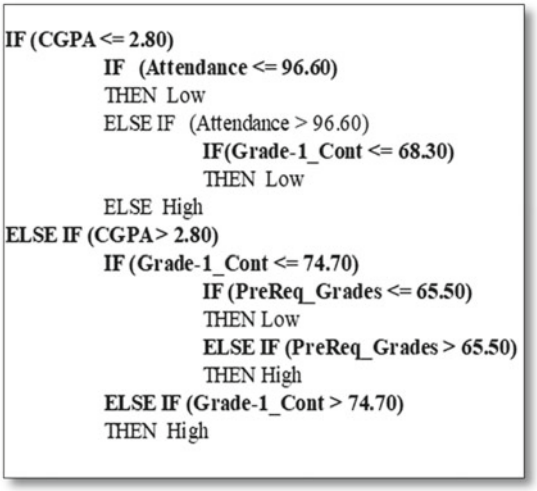


Fig. 7 The revealed classification rules

4 Conclusion

A major problem an instructor experiences is the efficient monitoring of students' academic progress in a course. The moment the students with unsatisfactory academic progress are identified, the instructor can take preventive measures to offer additional support to struggling students. We preprocessed a dataset, from an educational institute, and developed a set of machine learning prediction models. The decision tree dominates and is then transformed into a shape easily comprehensible by the instructor. This endows the instructor do devise supportive and preventive measures.

References

1. Norris D, Baer L, Leonard J, Pugliese L, Lefrere P (2008) Action analytics: measuring and improving performance that matters in higher education. EDUCAUSE Rev 43(1):42
2. Gašević D, Dawson S, Siemens G (2015) Let's not forget: learning analytics are about learning. TechTrends 59(1):64–71
3. Greller W, Ebner M, Schön M (2014) Learning analytics: from theory to practice–data support for learning and teaching. In: International computer assisted assessment conference. Springer
4. Alabri A, Al-Khanjari Z, Jamoussi Y, Kraiem N (2019) Mining the students' chat conversations in a personalized e-learning environment. Int J Emerg Technol Learn (iJET). 14(23):98–124
5. Romanenko V, Tropin Y, Boychenko N, Goloha V (2019) Monitoring student performance using computer technology. Slobozhanskyi herald Sci Sport 7(2(70)):36–39

6. Khan I, Al Sadiri A, Ahmad AR, Jabeur N (2019) Tracking student performance in introductory programming by means of machine learning. In: 2019 4th MEC international conference on big data and smart city (ICBDSC). IEEE
7. Asif R, Merceron A, Ali SA, Haider NG (2017) Analyzing undergraduate students' performance using educational data mining. Comput Educ 113:177–194
8. Iatrellis O, Savvas IK, Fitsilis P, Gerogiannis VC (2020) A two-phase machine learning approach for predicting student outcomes. Educ Inf Technol 1–20
9. Yang T-Y, Brinton CG, Joe-Wong C, Chiang M (2017) Behavior-based grade prediction for MOOCs via time series neural networks. IEEE J Sel Topics Sig Process 11(5):716–728
10. Umer R, Susnjak T, Mathrani A, Suriadi S (2017) On predicting academic performance with process mining in learning analytics. J Res Innovative Teach Learn
11. Asogwa O, Oladugba A (2015) Of students academic performance rates using artificial neural networks (ANNs). Am J Appl Math Stat 3(4):151–155
12. Costa EB, Fonseca B, Santana MA, de Araújo FF, Rego J (2017) Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. Comput Hum Behav 73:247–256
13. Drachsler H, Greller W (2012) Confidence in learning analytics
14. Romero C, Ventura S (2020) Educational data mining and learning analytics: an updated survey. Wiley Interdisc Rev Data Mining Knowl Discovery 10(3):e1355
15. Das K, Behera RN (2017) A survey on machine learning: concept, algorithms and applications. Int J Innovative Res Comput Commun Eng 5(2):1301–1309
16. Maesya A, Hendiyanti T (2019) Forecasting student graduation with classification and regression tree (CART) algorithm. In: IOP conference series: materials science and engineering. IOP Publishing
17. Orong MY, Caroro RA, Durias GD, Cabrera JA, Lonzon H, Ricalde GT (2020) A predictive analytics approach in determining the predictors of student attrition in the higher education institutions in the Philippines. In: Proceedings of the 3rd international conference on software engineering and information management
18. Chen H (2018) Predicting student performance using data from an auto-grading system. University of Waterloo
19. Saa AA (2016) Educational data mining & students' performance prediction. Int J Adv Comput Sci Appl 7(5):212–220
20. Pandey M, Taruna S (2016) Towards the integration of multiple classifier pertaining to the student's performance prediction. Perspect Sci 8:364–366
21. Alfere SS, Maghari AY (2018) Prediction of student's performance using modified KNN classifiers. Prediction of student's performance using modified KNN classifiers
22. Mondal A, Mukherjee J (2018) An approach to predict a student's academic performance using recurrent neural network (RNN). Int J Comput Appl 181(6):1–5
23. Lagman AC, Calleja JQ, Fernando CG, Gonzales JG, Legaspi JB, Ortega JHJC, Ramos RF, Solomo MVS, Santos RC (2019) Embedding naïve Bayes algorithm data model in predicting student graduation. In: Proceedings of the 3rd international conference on telecommunications and communication engineering
24. Jishan ST, Rashu RI, Haque N, Rahman RM (2015) Improving accuracy of students' final grade prediction model using optimal equal width binning and synthetic minority over-sampling technique. Decis Anal 2(1):1
25. Liao SN, Zingaro D, Thai K, Alvarado C, Griswold WG, Porter L (2019) A robust machine learning technique to predict low-performing students. ACM transactions on computing education (TOCE). 19(3):1–19
26. Ma X, Yang Y, Zhou Z (2018) Using machine learning algorithm to predict student pass rates in online education. In: Proceedings of the 3rd international conference on multimedia systems and signal processing
27. Rahman MHAA, Mustapha A, Razali N, Fauzi R (2018) Bayesian approach to classification of football match outcome. Int J Integr Eng 10(6)

28. Uqaili I-U-N, Ahsan SN (2020) Machine learning based prediction of complex bugs in source code. Int Arab J Inf Technol 17(1):26–37
29. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The WEKA data mining software: an update. ACM SIGKDD Explor Newslett 11(1):10–18
30. Hastie T, Tibshirani R, Friedman J, Franklin J (2005) The elements of statistical learning: data mining, inference and prediction. Math Intell 27(2):83–85
31. Matthews BW (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochimica et Biophysica Acta (BBA)-Protein Struct 405(2):442–451

A Review on Intrusion Detection System Using a Machine Learning Algorithms

Huda Sirag and Sally Dafaallah Awadelkariem

Abstract Intrusion detection has achievement a comprehensive interest and become a substantial scope for various researches and still being the subject of common concern by researchers. The intrusion detection researches face a complicated problem even after many years of research. Because of the huge number of false warnings throughout detecting anonymous attack patterns residue a vague problem. There are a numerous research trying to expose the possible solutions to this problem. Anomaly detection is the main issue of intrusion detection in which alarms of normal actions show a presence of destined or unintentional prompted attacks, faults, defects, and others. These papers represent an attentive literature review of machine learning approaches for intrusion detection, specifically Random Forest and Support Vector Machines that are used. Founded on the reference number and the significance of the growing method, papers describe every method specified, read, and shortened.

Keywords Intrusion detection system · Machine learning · Random forest · Support vector machines SVM

1 Introduction

Machine learning is a type of artificial intelligence which goals to enable machines that perform their tasks accurately and successfully through using intelligent software. The statistical learning approaches create the fundamentals for intelligent software which applies to advanced machine intelligence. As machine learning algorithms need data to obtain, the discipline must have a link through the discipline of the database. Machine learning integrates with additional concerns about what computing models and algorithms are often won't to more efficiently catch,

H. Sirag
Princeton University, Princeton, NJ 08544, USA

S. D. Awadelkariem (✉)
Springer Heidelberg, Tiergartenstr. 17, 69121 Heidelberg, Germany

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_24

collection, index, regain and combine these data, how multiple learning subtasks often orchestrated during a superior structure, and questions of computing manageability. Machine learning is an algorithm that will learn from data unaccompanied by counting on guidelines-based programming and numerical modeling is the validation of contacts among variables within the sort of mathematical equations [1]. Also, a machine learning models train to categorize based on input and output structures by a pre-defined number of classes. If a feature or distribution modification over time, then machine learning models will cause a substantial degradation in their performance. Though, if these models keep based on incoming data, then they cannot retain knowledge of the recurrent context [2]. In general machine learning algorithms can be classified as in the following subsections.

A. **Supervised Learning Algorithm**

This algorithm contains a goal variable that expected from a particular set of predictors. Via these sets of variables, which produce functions that map inputs to choose outputs. The training procedure remains until the model reaches a required level of precision on the training data. The main aim of supervised learning is measuring the range of data and create expectations of inaccessible, upcoming, or hidden data created on the labeled data model. Applied supervised algorithms are linear regression; logistical regression; random forest; support vector machines (SVM); nearest neighbor; decision trees; Naive Bayes; neural networks [3]. Supervised machine learning contains two primary methods: classification and regression.

Classification is a technique that received data and categorized it based on previous data samples then automatically trains the algorithm in order to identify and determined the objects types and classify them consequently. The system consumes learned with aids to certain types of information, achieve a visual character, image, or binary identifier [3].

Regression is procedure from classifying forms and computing the expectations of nonstop results. System recognizes the numbers, values, grouping, etc. The unsupervised algorithm does not contain any target variable to predict in another word, there is no direct control of the designer and defines information by examining through it and creation logic of it. Chosen results are unidentified and so far to be distinct. The furthermost broadly used algorithms are association rule: k-means clustering and principal component analysis (PCA); [4].

B. **Unsupervised Learning Algorithm**

Unsupervised learning algorithms used subsequent methods to define the data follows:

- Clustering: it is a survey of data that applies to the sector for meaningful groups established on their interior patterns deprived of previous knowledge of group identifications. The identifications are well-definite by the correspondence of separate data items and also parts of its difference from the remainder (that can apply to perceive anomalies).

- Dimensionality reduction: is tons of noise inside the received data. Machine learning algorithms usage dimensionality reduction for getting rid of the noise though filtering the significant information [5].

C. Reinforcement Learning Algorithm

This algorithm is an expert algorithm and definite decisions in addition trains itself frequently by a probationary plus error during this algorithm machine learns from previous experience and take the simplest potential knowledge to custom precise decision [5]. Reinforced use a method named exploration/exploitation. The action takings place when the results are experiential, so the following action reflects at primary achievement of the result. And inside the middle of the algorithms there are returning signals that happen upon execution of exact tasks. The purpose of the compensation signal might differ counting on the knowledge character. Therefore, reward signals could also be more classified as counting the desired operation. The system attempts to exploit positive rewards and decrease the negatives ones. The famous algorithms comprise Q-Learning; Temporal Difference (TD); Asynchronous Actor-Critic Agents (A3C) and Monte-Carlo Tree Search (MCTS) [6].

Reinforcement machine learning suit for many occasions of varying information are available. On this occasion, an algorithm custom processes based on connections between data and related procedures. Recent NPCs plus video games usage, this sort of ML. Reinforcement machine learning offers suppleness for AI responses to user's acts, therefore presuming sustainable defies. Self-driving also banks are sample of reinforced machine learning. For instance, when a self-driving car senses the road, when trying to turn to the left—the "turn left" state might start [7].

2 Intrusion Detection System

Debit to the supper use of the Internet, there are massive numbers of devices. Because of this enormous mass of network and the unidentified construction of Internet, protective equally communication plus information has become serious issues. Most of the systems use prevention techniques such as firewall, Intrusion detection systems (IDSs), a vital part to raise the system security level security. Attackers are always discovering novel ways to escape the avoidance system mechanism. IDSs come to be a convincing component of security systems [8]. IDSs units have three major components as presented in Fig. 1.

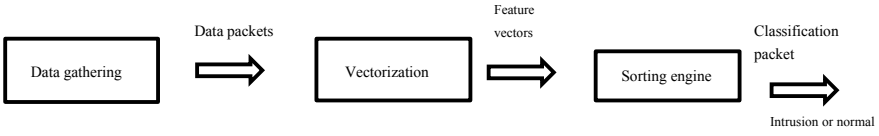


Fig. 1 Intrusion detection system Main component

The first component is a data-gathering mechanism that follows the currents of the network, formerly, these data requirements are used to recognizing the attribute and attribute vector desires to form. At last, this vector used, and it performs the classification engine [9]. The pivotal portion of IDS is the classification engine unit that selects if the improved feature vector is an intrusion for executing a dynamic intrusion detection system, primarily the irregularity detection techniques required, especially with a machine learning mechanism intrusion detection dataset as contain a negative influence to the accuracy of IDS [10, 11].

IDS divided according to their action, one type is an active-famous as an intrusion detection and protection system. It produces warnings and logs passes counting with instructions to adjust the structure to keep the network. The second type is passive-identifying hateful activity and produces a warning or logs; it doesn't takings any act. The former detects identified attacks with a high precision through pattern similar on known signatures cannot identify new attacks, as their signatures are not existing for pattern similarity [12]. IDS using many machine learning techniques, such as: K-means, Bayes Network, Random Forest Classifier, and Support Vector Machines algorithm [13].

The aim of this survey is to categorize and review the machine learning based on IDSs submit to date, abstract the most concepts of distributed machine learning to security field difficulties. This survey mainly planned to present a random forest algorithm and a survey of various techniques proposed by several researchers. Various machine learning algorithms useful to IDSs, where maybe supportive of machine learning research. Data objects are the foremost essential components in IDS. Data objects take features associated with attack behaviors. Feature types and have removal methods vary among distinct data elements, producing the most suitable machine learning models to vary too.

3 Random Forest Algorithm

Random Forest (RF) algorithm is a collaborative supervised machine learning algorithm. The classification performance of this algorithm is superior, in additional to single classifier models in furthermost cases and it can process equally binary and multi-classification. Procedure arbitrarily is the difficulty of the central concept of RF. When a choice a tree made, the info set must divide into sub-trees, supported by the simplest combination of variables. However, it's not easy to seek the proper combination of variables. A RF made from aggregate results of this forest of assembled trees. RF makes better results than separate decision trees. RF algorithm employed for equal classification and regression. The RF also handles lost value errors. The algorithm works in two stages first is formed and within the second RF, and a classifier prediction are made [14]. RF algorithm used in [15] to identify four types of attacks that are denial-of-service (DOS), Probe, Remote to Local (R2L), and User to Root (U2R) by dropping the features. The study assumed cross-validation to check the correctness of the applied algorithm for classification.

Based on feature selection measures where the first to classify various features according to attack types. Then, selected the best features for these attacks and, after that, applying RF and decision tree method to detect intrusion. In the final step amount the f-measure, precision, recall and accuracy, were measured and examine the result. The principal purpose of the feature selection procedure on the KDD'99 datasets which consists of 42 attributes is to decrease dimensionality, to eliminate redundancy of data, and deletion of irrelevant features. ANOVA F-test applied for unvaried features selection. All experiments prepared in the Python IDE named Anaconda. The attained result of RF classification with reduced features is more precise than which founding from all features.

In [16] RF algorithm is proposed to sense four kinds of an attacks R2L, probe, U2R, and DOS. They adopted and applied ten cross-validations to classification. The model of a pre-processing technique is applying discretization after loading the data set. Then, 4 datasets were grouped and dividing into training and testing. After that selected features of the superlative set utilize the subset feature choice measure by symmetrical uncertainty indemnify information gain as Eq. (1) [16].

$$SU(X, Y) = 2\left[IG\frac{X/Y}{H(X)H(Y)}\right] \tag{1}$$

Data set is specific to the RF for training and the tested data set served to the RF for sorting, finally precision, recognition rate, false warning rate and Mathew correlation measurement were computed for the empirical analysis. The NSL-KDD dataset was used in ARFF design. The subsequent preprocessing methods is used for running the experiment. (1) Substitute absent values: In weka, the researchers applied to substitute misplaced values filter to substitute all absent feature values in the NSL-KDD. The filter substitutes all absent values through the mean and manner of training data. (2) Discretization: Numeric features were discretized using discretization filter by unsupervised ten bin discretization. The final result demonstrates that precision, MCC plus DR, and of 4 kinds of attacks are improved after using the proposed method.

In [17], the KDDCup99 test datasets are examined via definite machine learning algorithms to define the precision of these algorithms by categorizing attacks to their several classes. It implements a practical study method during the study. The experimental results show that, the RF and random tree algorithms perform the best proficient in execution on the test dataset for sorting method. The experimental instrument by WEKA is used to achieve an association-based feature selection on the dataset with a best first search process, and the parameters applied for the calculation, recall, F-measure, and precision.

RF characteristic selection algorithm and multi-classifier collective applied based on deep learning for IDS in [18] to abstract ideal feature subset which is used to train by support vector machine, and k-nearest neighbor classification algorithm decision tree, after that, put in the deep learning to stack the output of four classifiers. Area under the curve (AUC) applied based on permutation variable importance measure VI_{xj}^{AUC} AUC (VIM) for feature selection. An AUC-based

permutation VIM mainly indicates the importance of the variable by calculating the average reduction before and after the area under the curve after the slight disturbance of the data independent variable. Therefore, the variable importance measure AUC of xj predictor calculated as shown in Eq. (2) [18]:

$$VI_{xj}^{AUC} = \frac{1}{ntree}\sum_{t=1}^{ntree}(AUC_{tj} - AUC_{\hat{tj}}) \tag{2}$$

where, ntree indicates the trees in the forest,AUC_{tj} indicates the AUC calculated of the OOB interpretations before permuting predictor xj in tree t, $AUC_{\hat{tj}}$indicates the AUC calculated from the OOB interpretations afterward permuting predictor xj randomly in tree t. they propose a multi-classifier ensemble method based on deep learning with k-foldcross-validation technology the experimental results illustrate that through the RF feature selection algorithm to abstract optimal feature subset, can efficiently improve the precision of intrusion detection.

In order to augment the recognition concert of IDS, [19] present a mixture of data adjusting technique created on heightening ML algorithms. The data adjusting technique contains two portions: data sampling in addition to feature selection. (i) Data sampling: iForest outlier recognition technique in order to sample data, a genetic algorithm applies to enhance the sampling ratio universally, and applied the sorting performance of RF on applicant sampled data as the estimation indicator. The major goal of data sampling is to examine the ideal training dataset and decrease the inequity for the data set. Other part is (ii) Feature Selection: applied the collaborating genetic algorithm with RF chooses features. Persistence of feature selection is to discover the greatest feature subset that can exploit the act of detection. When the peak training dataset and the peak feature subsection nominated, they fed at the classifier phase that occupations the RF then fixed suitability function as the classifier precision. In this study, F1 score supposed the fitness function, which is a consistent function that earnings precision and recalls. Computing the F1 score described as Eqs. (3), (4) and (5), respectively [19].

$$\text{F1 } score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{3}$$

$$Precision = \frac{\text{F1 } score}{\text{TP} + \text{FP}} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

where TP is referring to True Positive which is the real irregular records classified amount as irregular ones, in the other hand True Negative (TN) is the amount of definite regular accounts assorted as normal ones; False Positive (FP) is the amount of real normal accounts assorted as irregular ones, plus False Negative (FN) considered the amount of real irregular accounts categorized as normal ones. By equating with customary machine learning approaches, it shows that RF

classifier is a greatly robust classifier, thus the united outcome of data optimization and RF classifier creates DO IDS nearly continuously the superior between all the especially in distinguishing the anomalous performances through minimal records, for instance, DoS, Backdoor, and Analysis.

A new online lightweight learning technique AE-IDS proposed in [20] that use RF to choose the operative features from unique datasets. Online recognition straightly usages the designated features. It achieved influential illustration and decreases dimensionality, which enhanced the output of neural networks and conventional ML algorithms, for example, GMM, Kmeans, in binary sorting. AP clustering algorithm groups the superlative features to discover the features that have a robust connection, for aids feature to integrate. As the arrangement of feature amount altered, the upcoming results resort for biased to features with a superior arrangement of scale, causing in inferior AE expression capability. Clustering operations extremely associate features to a similar set that comprises features of similar scale arrangement. The renovation of their technique lies in the mixture of a three-layer narrow AE neural network plus a conventional unsupervised machine learning clustering algorithm. The AE-IDS uses a shallow AE neural network to decrease calculation difficulty plus accomplish the lightweight persistence. Furthermore, the precision of the unsupervised clustering algorithm has the ability to enhance by character-learning three-layer auto-encoder.

The work of [21] Concern the execution of LSTM with two techniques: a homogeneous technique—by enhanced bagged RF algorithm, and a heterogeneous technique—an Averaged Probability method of voting. The heterogeneous established on Decision Tree, Naïve Bayes, RIPPER, and kNN, with different computational features. The model applied on the UNSW_NB15 datasets in a 2-classed attack and as a multi-attack. LSTM attained a detection precision amount of 80% on the two-classed attack dataset and 72% precision amount in the multi-attack dataset. The homogeneous technique had an accuracy rate superior to heterogeneous model of 98%, 87.4% for two-class attack and therefor the multi-attack, correspondingly.

4 Support Vector Machines Algorithm

Is a linear ML technique founded in the structural risk decreasing of numerical learning theory (STL). SVM expressions respectable generality address, without the requisite for prior understanding. It usages a feature called kernel function for this mapping. Kernel functions similar to polynomial, radial basis functions are utilized for divide the feature space by creating a hyperplane. The kernel functions applied at the time of preparation of the classifiers, which chooses SVM beside the shallow of this function. SVM sort data by these support vectors that summarize the hyperplane in the feature space. It does not experience from the local lowest and its holder piercing datasets. These features parallel with the desires required to applying effective IDS Binary classification problems; resolved using SVM. An SVM maps linear algorithms into non-linear space [22]. In [23] constructed two

models for the sorting goals. The first built on SVM, while the second is RF. Experimental results illustration that both classifiers are operative. SVM is faintly additional precise extra affluent in expressions of time. RF products parallel precision in a considerably faster way uncertainty set modeling factors. These classifiers participate in an IDS system by a single source of analysis and raise its precision. KDD'99 Dataset applied to detect that one is the superlative detector for the dataset. Statistical study on the KDD'99 dataset established significant issues that extremely influence the concert of assessed systems and upcoming results in a reduced assessment for anomaly detection methods. The furthermost vital shortage in KDD'99 is vast number of dismissed records. To resolve these matters, they established a new dataset, KDD99Trai, and KDD99Test, that does not contain any dismissed records in the train set along with the test set, consequently the classifiers will not influence near regular records. The result is more convenient to apply RF and SVM expressively, so that makes the most of the performance percentage and decreases the incorrect-negative rate.

In this study [12], a new technique submitted to distinguish DDoS attacks established on the RF and SVM algorithm. By considering the value of attribute choice in DDoS attacks recognition, the RF-SVM algorithm destined to achieve RF to calculate the feature value and SVM to re-screen the structures that will avoid eliminating structures incorrectly. The study picks ID3 as the algorithm of calculating feature value, that standard in the information gain. Entropy denotes a disordered state. The information entropy is definite in Eq. (6) as [12]:

$$\text{Entropy}(\text{x}) = \sum_{i=1}^{c} p_i \log_2 p_i \tag{6}$$

The arbitrary variable X is a fixed arbitrary discrete variable, Pi is the proportion of arbitrary variable xi and the dataset C. Uncertainty the training model x distributed by structure Y, the probability of Y for x shown in Eq. (7) [12]:

$$\text{Entropy}_Y(x) = \sum_{J=1}^{N} \frac{|x_i|}{|x|} \text{Entropy}_Y(x) \tag{7}$$

While the improbability of training samples increases, the information entropy rises too. The description of information gain described as follows in Eq. (8) [12]:

$$Gain(x, Y) = \text{Entropy}(\text{x}) - \text{Entropy}_Y(x) \tag{8}$$

Lastly, an ideal feature subset attained, that reached the advanced recognition rate and requisition rate. Two types of datasets used in this study to train and test. The result demonstrates that the RDF-SVM algorithm can choose the ideal feature subset above the KDD99 dataset and can also differentiate among the attacks traffic by DDoS with standard traffic above the DDoS dataset composed from the actual environment. Likened through the Logistic Regression, Neural Network, AdaBoost,

and SVM technique, the RDF with the SVM algorithm has an advanced recognition rate and requisition rate [12].

RF and SVM presented in [24]. For exercise, the model used various steps that contain data collection, preprocessing, feature selection, a model for training in addition to training, and confirmation. The study has shown a comprehensive investigation of the two ML algorithms consuming the whole features; it detected for time-consuming and performance debasing. The study investigates feature selection behaviors that are used to determine subjective causes and eliminate redundant features of data that ensure not disturb the precision of projecting models. The crucial difference among feature selection besides abstraction is that feature choice attempts to catch out on the finest subset of features through unique features, while feature extraction makes a set of new features. As the experience shows, a limited of the features in the dataset dismissed and unconnected. Feature choice also shows a necessary part of the suggested work. Recursive Feature Elimination organized to decrease the dimensionality of the dataset. In the contrast, RF achieved superior to SVM earlier feature selection. SVM achieved superior to RF after feature choice for the furthermost of attacks.

5 Conclusion

Machine learning techniques have got extensive attention among the intrusion detection researchers to study the weaknesses of knowledge-based detection techniques. This paper first proposes a machine learning algorithm type. Followed by discussing intrusion detection system IDSs taxonomy. This paper defines a literature review for Machine learning ML approaches used for the intrusion detection system. Summarizing the recent studies that used Random Forest and Support Vector Machines to provide references to other researchers conducting in-depth studies.

References

1. Ngu AH, Gutierrez M, Metsis V, Nepal S, Sheng QZ (2017) Iot middleware: a survey on issues and enabling technologies. IEEE Internet Things J 4(1):1–20
2. Jameel SM, Hashmani MA, Alhussain H, Budiman A, Rehman M (2020) A critical review on adverse effects of concept drift over machine learning classification models. (IJACSA) Int J Adv Comput Sci Appl 11(1)
3. Shibahara T, Yagi T, Akiyama M, Chiba D, Yada T (2016) Efficient dynamic malware analysis based on network behavior using deep learning. In: Global communications conference (GLOBECOM). IEEE, pp 1–7
4. Karatas G (2016) Genetic algorithm for intrusion detection system. In: 24th signal processing and communication application conference (SIU). IEEE, pp 1341–1344
5. Imtiaz W, Khanum M, Mahboob T, Ghafoor HA, Sehar R (2015) A survey on unsupervised machine learning algorithms for automation, classification and maintenance. Int J Comput Appl (0975–8887) 119(13)

6. Kang M-J, Kang J-W (2016) Intrusion detection system using deep neural network for in-vehicle network security. PLoS ONE 11(6):e0155781
7. Tang TA, Mhamdi L, McLernon D, Zaidi SAR, Ghogho M (2016) Deep learning approach for network intrusion detection in software defined networking. IEEE conference location: Fez, Morocco date of conference. https://doi.org/10.1109/WINCOM.2016.7777224, 26–29 Oct 2016
8. Buczak A (2016) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Sur Tutorials 18(2)
9. Niyaz Q, Sun W, Javaid AY, Alam M (2016) A deep learning approach for network intrusion detection system. College of Engineering University of Toledo BICT 2015, December 03–05, New York City, United States Copyright © ICST. https://doi.org/10.4108/eai.3-12-2015.2262516
10. Karatas G, Demir O, Sahingoz OK (2018) Deep learning in intrusion detection systems. International congress on big data, deep learning and fighting cyber terrorism, Ankara, Turkey, 3–4 Dec 2018
11. Harmony A (2020) Search-based feature selection technique for cloud intrusion detection Widad Mirghani Makki, Maheyzah MD. Siraj, and Nurudeen Mahmud Ibrahim(&) Advances in intelligent systems and computing volume 1073© Springer Nature Switzerland AG 2020
12. Wang C, Zheng J, Li X (2017) Research on DDoS attacks detection based on RDF-SVM. In: 10th international conference on intelligent computation technology and automation
13. Balogun A (2015) Anomaly intrusion detection using an hybrid of decision tree and K-nearest neighbor a multidisciplinary journal publication of the Faculty of Science, Adeleke University, Ede, Nigeria
14. Yao H, Fu D, Zhang P, Li M, Liu Y (2019) MSML: a novel multilevel semi supervised machine learning framework for intrusion detection system. IEEE Internet Things J 6 (2):1949–1959
15. Kumar S, Pal J, Giri A, Raj A, Raj R, Mangayarkarasi R (2019) Intrusion detection system using random forest. researchgate.net/publication/332396674
16. Farnaaz N, Jabbar MA (2016) Random forest modeling for network intrusion detection system. Twelfth international multi-conference on information processing-2016 (IMCIP-2016)
17. Ugochukwu CJ, Bennett EO (2018) An intrusion detection system using machine learning algorithm. Int J Comput Sci Math Theory 4(1). ISSN: 2545-5699
18. Ling J, Wu C (2019) Feature selection and deep learning based approach for network intrusion detection. Advances in computer science research, volume 873rd international conference on mechatronics engineering and information technology (ICMEIT 2019)
19. Ren J, Guo J, Qian W, Yuan H, Hao X, Jingjing H (2019) Building an effective intrusion detection system by using hybrid data optimization based on machine learning algorithms. Secur Commun Netw I7130868:11
20. Li X, Chen W, Zhang Q, Wu L (2020) Building auto-encoder intrusion detection system based on random forest feature selection. Comput Secur 95. 101851xzzaza HM
21. Elijah AV, Abdullah A, JhanJhi NZ, Supramaniam M, Abdullateef B (2019) Ensemble and deep-learning methods for two-class and multi-attack anomaly intrusion detection: an empirical study. (IJACSA) Int J Adv Comput Sci Appl 10(9)
22. Enache A-C, Patriciu VV (2014) Intrusions detection based on support vector machine optimized with swarm intelligence. In: 153–9th IEEE international symposium on applied computational intelligence and informatics, Timişoara, Romnia, May 15–17
23. Al Mehedi Hasan M, Nasser M, Pal B, Ahmad S (2014) Support vector machine and random forest modeling for intrusion detection system (IDS). J Intell Learn Syst Appl 6:45–52 Published Online February 2014
24. Patgiri R, Varshney U, Akutota T, Kunde R (2018) An investigation on intrusion detection system using machine learning. IEEE symposium series on computational intelligence SSCI

A Systematic Review for Sentiment Analysis of Arabic Dialect Texts Researches

Arwa A. Al Shamsi and Sherief Abdallah

Abstract People in the Arab world prefer to use their dialects in writing while using social media sites which results in generating a significant amount of texts. Sentiment analysis in texts is one of the most important applications of the Natural Language Processing field of science. Sentiment Analysis involves classifying texts based upon emotions or polarity. Different approaches have been utilized so far for Sentiment Analysis of Arabic Texts. In this systematic review, the author aims to explore the researches and studies that were in the field of Sentiment Analysis of Arabic Dialects to report the utilized approaches, constructed datasets, most common type of dialects explored by researchers, and reported results of Sentiment Analysis in term of accuracy, recall, precision, and F-Measure. Author reported the following findings for this systematic review: (1) in most of the included research papers the dialect type is not specified and mostly it was mentioned (Arabic Dialects) in general, however, some other dialects were explored specifically by authors such as Saudi Dialect (2) 50% of the utilized datasets in the included research papers are constructed by the authors themselves, moreover, the size of most of the utilized datasets are between 10,000 and 50,000, and very limit number of datasets had size above 100,000, (3) machine learning algorithms (classifiers) are the most common approach that were used for Sentiment Analysis of Arabic Dialects (4) the best result for Sentiment Analysis of Arabic Dialects achieved while different machine learning algorithms (classifiers) were used, and (5) among social media platforms, Twitter is the most common utilized online platform for constructing datasets for Arabic Dialects texts.

Keywords Natural language processing · Arabic dialects · Sentiment analysis · Systematic review

A. A. Al Shamsi (✉) · S. Abdallah
Faculty of Engineering and IT, The British University in Dubai, 345015 Dubai, United Arab Emirates
e-mail: 20180935@student.buid.ac.ae

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_25

1 Introduction

Social Media has become a part of daily life; It is increasingly hard to survive in this modern era of digital media without a digital footprint [1]. People are increasingly using social media platforms for various purposes and reasons, resulting in a vast amount of online data being created on a daily basis [2]. Arabic is a Semitic most commonly utilized language around the world, it is the official language in 28 countries [3]. The Arabic Language has three forms i.e. Classical Arabic, Modern Standard Arabic, and Arabic Dialects. Dialects are a type of language that is spoken informally. Each Arab country has its own dialect, and each dialect has numerous sub-dialects [4]. In the Arab world, people mostly tend to use their dialects in writing while using social media sites [5]. Authors in [6] stated that there is a great need to process the online data that is in the form of natural language. The research community witnessing an increased interest in researches that are targeting Arabic Dialects texts such as text classification, identification, translation, and Sentiment Analysis. Sentiment Analysis or Opinion Mining involves the classification of texts based on emotions or opinions [7]. Sentiment Analysis is used to analyze people's opinions and feelings about organizations, services provided, and products. Different approaches have been used for Sentiment Analysis of Arabic Dialect texts. In this systematic review, the author aim to (1) finds the most common dialects explored in Sentiment Analysis researches, (2) the size of the used datasets, (3) the approaches that are used for Sentiment Analysis of Arabic Dialects texts, (4) the Evaluation Metrics that are used in evaluating the approaches for Sentiment Analysis of Arabic Dialects, and (5) the online platforms that are used in constructing datasets for Sentiment Analysis of Arabic Dialects researches. This paper organized as follows: Sect. 2 represents some of the Arabic Language characteristics and the related works and researches that explored Sentiment Analysis for Arabic Dialect texts, Sect. 3 gives description of the research methods used in this systematic review, Sect. 4 provides the results and the discussion, Sect. 5 presents the conclusion, last section incudes the references.

2 Related Works

There had been a great interest in the classification and analysis of Arabic Language texts. The Arabic Language has its unique features. Different approaches were used for the classification and Analysis of Arabic texts, moreover, different tools were utilized for the same purpose. The Modern Standard form of Arabic texts was the target by huge researches earlier, however, recently increasing researches targeted the Arabic Dialects for classification and analysis. In this section, the author presented some of the Arabic Language characteristics and the related works and researches that explored Sentiment Analysis for Arabic Dialect texts.

2.1 Arabic Language Background

The Arabic Language has been used for about 2000 years in the Arabian Peninsula and it was in the form of Classical Arabic which is the form of Arabic in the Holy Quran [8]. The Arabic Language is a popular language that is spoken by a huge number of people around the world. Reference [9] mentioned that the Arabic language considered as the fifth most common language all around the world. The Arabic Language has some special characteristics [10]. The Arabic language is considered a Semitic language as it is written from the right side to the left side. Moreover, the authors mentioned that the shape of the Arabic letters changed depending on their position in the word itself [7]. The Arabic Language consists of 28 letters with no upper case and lower case distinguish, numbers can be presented as singular, dual, or plural.

2.2 Arabic Natural Language Processing ANLP

ANLP involves enabling computers to analyze Arabic Natural Language by developing different tools and techniques for the Arabic language that is in either text form or human speech form [7]. Although it is expected that ANLP tools can be utilized for all the forms of the Arabic language, ANLP is rarely used for Classical Arabic, moreover, most of the existing tools of ANLP were designed for the Modern Standard form of Arabic language compared to the Arabic Dialectal form. Authors in [11] stated that it is hard to develop ANLP tools for Arabic dialects because of the complexity of Arabic Dialects compared to MSA such as the phonological and morphological challenges. ANLP researches and implementations recently involve building resources for Arabic texts, basic language analysis, Identification of the language, and Semantic language analysis such as Sentiment Analysis and Machine Translation. In this systematic review, the author is focusing on Sentiment Analysis of Arabic Dialects.

2.3 Sentiment Analysis of Arabic Dialects

Sentiment Analysis involves the analysis and classification of texts based on emotions or opinions. Sentiment Analysis aims to analyze people's opinions about organizations, provided services, or products, moreover, it is important to analyze sentiments in social media sites as it would provide valuable information [7]. Different approaches have been used so far for the analysis of the sentiment in Arabic Dialect texts. The most common approaches are Machine Learning algorithms (classifiers), Deep Learning models, a Lexicon-Based approach, and a hybrid approach that combined both machine learning algorithms and a Lexicon-based

approach. In [12–24] authors used machine learning algorithms (classifiers) approach for Sentiment Analysis of Arabic Dialects, and the most common used algorithm is SVM followed by NB. In [10, 25–27], authors utilized lexicon-based approach. Machine learning algorithms are widely used for Sentiment Analysis of Arabic Dialects compared to the lexicon-based approach. While in [28–30] hybrid approach was utilized for Sentiment Analysis of Arabic Dialects in which both machine learning algorithms and lexicon-based approach were used. Deep Learning models as well increasingly utilized recently for Sentiment Analysis of Arabic Dialects as shown in [3, 31–37]. Moreover, the authors of [38, 39] used Deep Learning models and Machine Learning algorithms (classifiers) to compare the Sentiment classification results.

2.4 Sentiment Analysis of Arabic Dialects Challenges

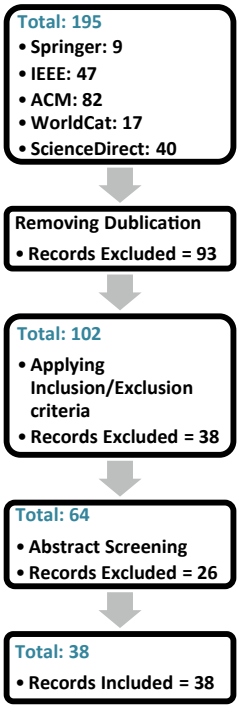
ANLP researches in the field of Sentiment Analysis that targeted Arabic Dialects are increasing, however, several challenges associated with Sentiment Analysis of Arabic Dialects raised. Authors in [40] stated that the Arabic language is known for its complexity, moreover, researches and studies that have been conducted in the ANLP field are fewer than NLP for English texts. Additionally, [11, 41] stated that Arabic Dialects presented an orthographic challenge as it does not have a standard written form for the texts, this challenge has led to a lack of NLP tools for Arabic Dialects. Authors of [42] stated that diacritical may change the meaning of the same word, negation is complicated compared to the English language, and Arabic Dialects does not have a standard written form which may result in spelling errors. From above, Sentiment Analysis of Arabic Dialects challenges are lack of tools for Arabic Dialects, Lack of resources i.e. datasets and lexicons, orthographic challenge, diacritical may change the word's meaning, negation is challenging and more complicated than negation in the English language.

3 Research Method

In this Systematic review, the author first Identified the research questions, then the author Identified the search strategy, after that the keywords that are used in the research were identified, then the quality assessment questions were identified. The author then extracts the research papers and performs a critical analysis for the extracted papers. Figure 1 illustrates the steps that are followed in this systematic review.

Fig. 1 The steps followed in this systematic review

Fig. 2 Data source



3.1 Data Source

The author started working on this study in November 2020, the author targeted research papers that are published in the period from 2015 to 2021. The databases the author utilized are IEEE, Springer, ScienceDirect, ACM, and WorldCat. Figure 2 illustrates the databases that are used and the number of research papers from each database.

3.2 Search Strategy

The keyword used to collect the research papers:

- "Arabic Dialects" and "Sentiment Analysis"
- "Arabic Dialects" and "Sentiment Analysis" and "deep learning"
- "Arabic Dialects" and "Sentiment Analysis" and "classification"

3.3 Inclusion/Exclusion Criteria

It is essential to clarify the inclusion and exclusion criteria to make sure that the research papers that will be included in this systematic review are relevant and would result in achieving the aim of the study. Note that most of the included papers are with open access, some papers whose access was restricted were excluded.

- Must involve Sentiment Analysis experiment study
- Must be for Arabic Dialect texts only
- Must present the Evaluation Metrics
- Must address information about the utilized corpus
- Must be published in a journal or conference

3.4 Quality Assessment

In this section, the author design quality assessment questions for checking whether the included researches satisfied the main goal of this systematic review.

Q1: Was the size of the used dataset identified? Is it mentioned whether it is open source or created by the author?

Q2: Are the sentiment analysis approaches used in the Sentiment Analysis experiment well identified and described?

Q3: Are the Evaluation Metrics that are used in evaluating the approaches for Sentiment Analysis well identified and described?

Q4: Are the results in terms of accuracy, recall, precision and F-measure reported?

Q5: Is the online platform that is used in constructing the dataset identified?

Note: In this assessment, a scale was used, a 3-point scale was used for Q1, Q2, Q3, and Q5, while for Q4 4-point scale was used. In 3-point scale for Q1, Q2, Q3, and Q5, if the included paper addressed the question completely, then it would be graded 1 point if it addressed the question partially then it would be graded (0.5) point, while if it did not address the question then it would be graded 0. In 4-point scale for Q4, if the included paper addressed the question completely, then it would be graded 1, if it addressed the question partially then it would be graded (0.25) for each evaluation metric used in that paper, while if it did not address the question

Table 1 Quality assessment evaluation for the included papers in systematic review

Ref.	Study	Q1	Q2	Q3	Q4	Q5	Total	Percentage (%)
[12]	S1	1	1	1	1	1	5	100
[25]	S2	1	1	1	0.25	1	4.5	85
[43]	S3	1	1	0.5	1	1	4.5	90
[31]	S4	1	1	1	0	1	4	80
[10]	S5	1	1	1	1	1	5	100
[32]	S6	0.5	1	1	1	0.5	4	80
[44]	S7	1	1	1	1	1	5	100
[13]	S8	1	1	1	1	1	5	100
[33]	S9	1	1	1	1	1	5	100
[45]	S10	1	1	1	1	1	5	100
[14]	S11	1	1	1	0.25	1	4.25	85
[27]	S12	1	1	1	0.5	1	4.5	90
[15]	S13	1	1	1	0.25	0.5	3.75	75
[38]	S14	1	1	1	0.25	1	4.25	85
[39]	S15	1	1	1	0.5	1	4.5	90
[16]	S16	1	1	1	0.25	1	4.25	85
[34]	S17	1	1	1	0.25	1	4.25	85
[17]	S18	1	1	1	1	1	5	100
[28]	S19	1	1	1	0.5	1	4.5	90
[18]	S20	1	1	1	1	1	5	100
[29]	S21	1	1	1	0.25	1	4.25	85
[46]	S22	1	1	1	0.25	1	4.25	85
[47]	S23	1	1	1	0.25	1	4.25	85
[35]	S24	1	1	1	0.25	1	4.25	85
[36]	S25	0.5	1	1	0.5	1	4	80
[19]	S26	1	1	1	0.25	1	4.25	85
[30]	S27	1	1	1	0.75	1	4.75	95
[20]	S28	1	1	1	1	1	5	100
[26]	S29	1	1	1	0.25	1	4.25	85
[37]	S30	1	1	1	0.5	1	4.5	90
[21]	S31	1	1	1	1	1	5	100
[22]	S32	1	1	1	0.5	1	4.5	90
[48]	S33	1	1	1	0.5	1	4.5	90
[49]	S34	1	1	1	1	1	5	100
[50]	S35	1	1	1	1	1	5	100
[23]	S36	1	1	1	0.75	1	4.75	95
[24]	S37	1	1	1	1	1	5	100
[3]	S38	1	1	1	0.75	1	4.75	95

then it would be graded 0. Table 1 illustrates the quality assessment evaluation for the included research papers.

4 Results and Discussion

In this section, the author discussed the results for each of the systematic review research questions. The findings are illustrated in tables and figures as presented below.

RQ1: What are the most common dialects that are considered in Sentiment Analysis experiments researches?

Table 2 and Fig. 3 presents the dialect types that were considered in the included research papers. It is found that in most of the papers, authors didn't specify the dialect type, and mostly it was mentioned (Arabic Dialects), and this may be due to the fact of difficulty to collect resources (datasets) for a specific dialect to be used for Sentiment Analysis purposes; as people from different regions may use the same online platform, and may write comments on the same post, and they may follow the same account or profile, hence, it is hard to filter the collected comments, posts, or tweets to represent specific dialect and build resources for a specific dialect. Moreover, Table 2 presents some important dialects that were explored specifically for Sentiment Analysis i.e. Saudi Dialect, Levantine Dialect, Jordanian Dialect, Algerian dialect, Tunisian Dialect, Egyptian Dialect as illustrated in Fig. 3.

RQ2: What is the size of the used dataset of Arabic Dialect that was used in the Sentiment Analysis Experiment? Is it open source or created by the author?

Table 3 illustrates that 50% of the utilized datasets in Sentiment Analysis of Arabic Dialects are constructed by the authors themselves. Table 3 and Fig. 4 present that authors of the included papers mostly utilized the Twitter platform for constructing the datasets, followed by websites (blogs, book reviews, movie reviews, etc.), followed by the Facebook platform. However, recently different social media platforms are widely used by Arab world populations, such as Instagram, YouTube, Linked In, etc. and these platforms have not been utilized so far for constructing resources for Arabic Dialects. An important finding reported in Table 3 and Fig. 5 that the size of most of the utilized datasets are between 10,000 and 50,000, and a very limited number of datasets whose size is above 100,000, taking into consideration that the great size of the dataset may positively affect the performance results such as accuracy results of the Sentiment Analysis experiment.

RQ3: What are the sentiment analysis approaches used in Sentiment Analysis of Arabic Dialects?

Table 4 illustrates the approaches that have been used for Sentiment Analysis in the included papers in this systematic review. It represents that the machine learning

Table 2 Arabic dialect types explored in the included papers in this systematic review

Ref.	Study	Dialect type	Ref.	Study	Dialect type
[12]	S1	MSA and Jordanian dialect	[18]	S20	MSA and Arabic dialects
[25]	S2	MSA and Algerian dialect	[29]	S21	MSA and Arabic dialects
[43]	S3	Egyptian dialects and some other dialects	[46]	S22	MSA and Arabic dialects
[31]	S4	Arabic dialects	[47]	S23	MSA and Saudi dialect
[10]	S5	MSA and Egyptian dialects	[35]	S24	MSA and Arabic dialects
[32]	S6	Algerian dialect	[36]	S25	MSA and Arabic dialects
[44]	S7	Arabic dialects	[19]	S26	MSA and Arabic dialects
[13]	S8	MSA and Jordanian dialect	[30]	S27	MSA and Jordanian dialect
[33]	S9	MSA and Arabic dialects	[20]	S28	Jordanian dialect
[45]	S10	MSA and Arabic dialects	[26]	S29	MSA and Saudi dialect
[14]	S11	MSA and Levantine dialects	[37]	S30	MSA and Arabic dialects
[27]	S12	Classical Arabic, MSA and Arabic dialects	[21]	S31	MSA and Arabic dialects
[15]	S13	MSA and Arabic dialects	[22]	S32	MSA and Tunisian dialect
[38]	S14	MSA and Arabic dialects	[48]	S33	MSA and Arabic dialects
[39]	S15	MSA and Arabic dialects	[49]	S34	MSA and Arabic dialects
[16]	S16	Saudi dialect	[50]	S35	MSA and Tunisian dialect
[34]	S17	MSA and Arabic dialects	[23]	S36	MSA and Jordanian dialect
[17]	S18	MSA and Arabic dialects	[24]	S37	Levantine dialects
[28]	S19	MSA and Arabic dialects	[3]	S38	MSA and Arabic dialects

algorithms (classifiers) are the most common method followed by Deep Learning methods. Moreover, it represents that a limited number of researchers used the Lexicon-Based approach and Hybrid approach that combines both machine learning approaches and Lexicon-based approach.

RQ4: What were the Evaluation Metrics that are used in evaluating the approaches for Sentiment Analysis of Arabic Dialects?

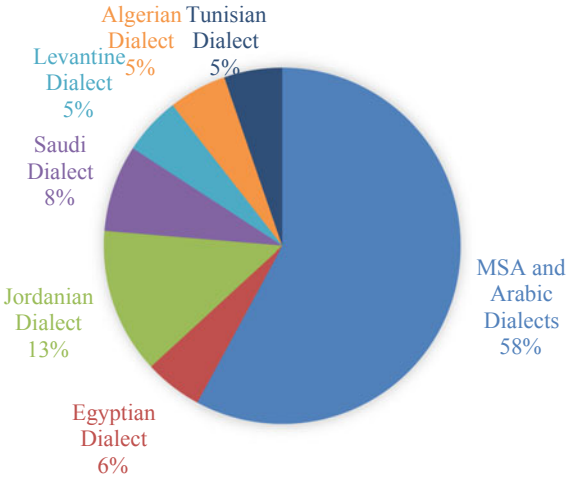


Fig. 3 Arabic dialects types explored in the included research papers

Table 5 illustrate that the most common Evaluation Metrics that were used are accuracy, followed by F-measures. This indicates that most researches aim to prove the accuracy of the developed framework for Sentiment Analysis. Moreover, represents that in most of the papers, either 1 or 4 Evaluation Metrics were used.

RQ5: What are the research results reported from the collected research papers in terms of accuracy, recall, precision, and F-measure?

Table 6 presents the results that were reported while implementing different approaches for Sentiment Analysis of Arabic Dialects. The best result was reported by [18] in which different machine learning algorithms (classifiers) were used for Sentiment Analysis of Arabic Dialects. Although machine learning algorithms (classifiers) were used widely for Sentiment Analysis of Arabic dialects, in [18] many factors impacted the implementation and improve the results, such as the use of a huge annotated dataset, the implementation of different preprocessing methods, and the utilization of n-gram features. On the other hand, the best result achieved while implementing the deep learning approach was reported in [33] in which authors used CNN and LSTM deep learning approaches and compared them with classical machine learning approaches (classifiers) and authors proofed that the deep learning model outperformed the machine learning algorithms (classifiers) approaches.

RQ6: Which online platforms are widely used in constructing datasets for the researches and studies for Sentiment Analysis of Arabic Dialects?

Table 7 illustrates that the most utilized platform for constructing datasets for Arabic Dialects texts is Twitter, followed by different websites (book reviews, hotel reviews, movie reviews, blogs, etc.), followed by the Facebook platform. Figure 4 presented the same finding. Twitter attracted researchers due to many reasons such as the tweets are formed in short sentences, Twitter API that makes it easy to export

Table 3 Datasets used in the included research papers in this systematic review

Ref.	Study	Corpus size	Platform	Annotated (Yes/No)	Constructed by author
[12]	S1	2400 reviews	JEERAN website	Yes	No
[25]	S2	7698 comments	Facebook	Yes	Yes
[43]	S3	8000 tweets	Twitter	Yes	Yes
[31]	S4	5600 tweets	Twitter	Yes	No
[10]	S5	2067 tweets	Twitter	Yes	No
[32]	S6	7287 comments	N/A	Yes	No
		10,254 comments	Facebook	Yes	
		10,366,630 sentences	N/A	Yes	
[44]	S7	1103 tweets	Twitter	Yes	Yes
[13]	S8	1800 tweets	Twitter	Yes	Yes
[33]	S9	2000 tweets	Twitter	Yes	Yes
[45]	S10	2071 reviews	Apple Store and Google Play	Yes	Yes
[14]	S11	2242 tweets	Twitter	Yes	Yes
		19,738 reviews	Book reviews		No
		13,158 reviews	Book reviews		No
		2899 tweets	Twitter		No
[27]	S12	15,274 reviews	Twitter, Facebook, public survey and mystery shopper	Yes	Yes
[15]	S13	8 datasets: 31,414 tweets	Twitter	Yes	No
[38]	S14	92,492 sentences from 8 public datasets	Twitter, book reviews, movies reviews, and blogs	Yes	No
[39]	S15	24,028 reviews	Hotel's reviews	Yes	No
[16]	S16	2010 tweets	Twitter	Yes	No
[34]	S17	2 datasets (10,689 tweets)	Twitter	Yes	No
[17]	S18	63,257 reviews	goodreads.com	Yes	No
[28]	S19	6 datasets: (38,977 tweets)	Twitter	Yes	No
[18]	S20	151,548 tweets	Twitter	Yes	Yes
[29]	S21	OCA: 500 reviews	Webpages and blogs	Yes	No
		ARMD: 1000 reviews	elcinema.com cairo360.com	Yes	Yes
[46]	S22	Main dataset: 2026	Twitter	Yes	Yes
		Sub dataset: 1732			
[47]	S23	8202 tweets	Twitter	Yes	Yes
[35]	S24	3 datasets: 89,091	Twitter	Yes	No
[36]	S25	N/A	Twitter	Yes	No

(continued)

Table 3 (continued)

Ref.	Study	Corpus size	Platform	Annotated (Yes/No)	Constructed by author
[19]	S26	151,500 tweets	Twitter	Yes	Yes
[30]	S27	22,550 tweets	Twitter	Yes	Yes
[20]	S28	1000 tweets	Twitter	Yes	Yes
[26]	S29	17,573 tweets	Twitter	Yes	Yes
[37]	S30	10,000 tweets	Twitter	Yes	No
[21]	S31	959 tweets	Twitter	Yes	Yes
[22]	S32	17,000 comments	Facebook	Yes	Yes
[48]	S33	7 datasets: 524,659	Social Media	Yes	No
[49]	S34	5 datasets: 26,956	Facebook and Twitter	Yes	No
[50]	S35	3 datasets: 16,297	Facebook and Twitter	Yes	No
[23]	S36	2730 reviews	JEERAN website	Yes	Yes
[24]	S37	1798 tweets	Twitter	Yes	Yes
[3]	S38	3 datasets: 32,260	Twitter	Yes	No

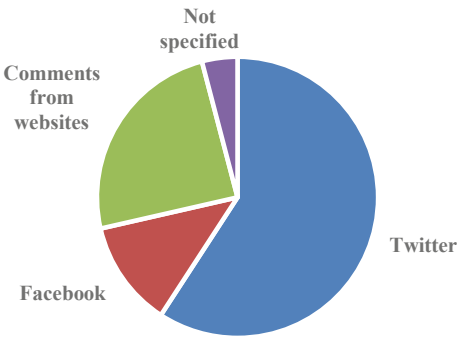


Fig. 4 Platforms used for dataset construction

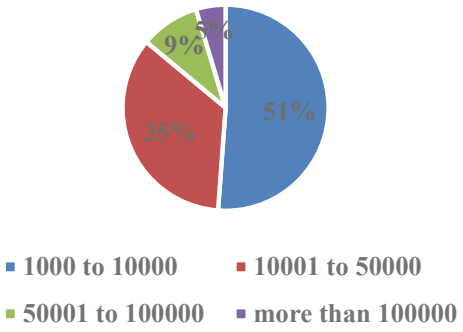


Fig. 5 Size of the used datasets

Table 4 Sentiment analysis approaches used in the included papers in this systematic review

Ref.	Study	Utilized approach for sentiment analysis
[12]	S1	SVM classifier
[25]	S2	Lexicon based approach
[43]	S3	Web-based tool (lexicon-based approaches or machine Learning approaches)
[31]	S4	Deep learning and ensemble implementations
[10]	S5	Hybrid lexicon approach (unsupervised and supervised technique)
[32]	S6	Deep learning
[44]	S7	Hybrid approach (machine learning and semantic orientation)
[13]	S8	SVM classifier and NB classifier
[33]	S9	Deep learning (CNN and LSTM)
[45]	S10	Aspect-based sentiment analysis
[14]	S11	Machine learning algorithms
[27]	S12	Hybrid model (corpus-based and lexicon-based models)
[15]	S13	Machine learning algorithms
[38]	S14	Machine learning algorithms and deep learning (CNN)
[39]	S15	Machine learning algorithms and deep learning (SVM and RNN)
[16]	S16	Machine learning algorithms (SVM, MNB, SGD, KNN, LR, PA)
[34]	S17	Deep learning (CNN)
[17]	S18	Machine learning algorithms (SVM, NB, DT, KNN)
[28]	S19	Hybrid approach (lexicon-based and machine learning)
[18]	S20	Machine learning algorithms
[29]	S21	Hybrid approach (lexicon-based and machine learning (SVM))
[46]	S22	Machine learning algorithms (BNB, MNB, NSVC, LSVC, SGD, RGD, LR)
[47]	S23	Machine learning algorithms (SVM, NB, KNN, LR, MLP)
[35]	S24	Deep learning (CNN and LSTM)
[36]	S25	Deep learning (narrow CNN)
[19]	S26	Machine learning algorithms (SVM, NB, BNB, MNB, SGD, LR)
[30]	S27	Hybrid approach (lexicon-based and machine learning (SVM and NB))
[20]	S28	Machine learning algorithms (SVM and NB)
[26]	S29	Hybrid model (corpus-based and lexicon-based models)
[37]	S30	Deep learning (CNN and LSTM)
[21]	S31	Machine learning algorithms (SVM, NB, and KNN)
[22]	S32	Machine learning algorithms (SVM, BNB, MLP)
[48]	S33	Shallow neural network (syntax-ignorant n-grams embeddings are composed and learned in a supervised manner)
[49]	S34	Shallow neural network (syntax-ignorant n-grams embeddings feature)
[50]	S35	Machine learning algorithms (SVM and NB) and Lexicon-based approach
[23]	S36	Machine learning algorithms (SVM)
[24]	S37	Machine learning algorithms
[3]	S38	Deep learning

Table 5 Evaluation metrics used for evaluation in the included papers

Ref.	Study	Evaluation metrics used for evaluation				
		Accuracy	Recall	Precision	F-measure	other
[12]	S1	✓	✓	✓	N/A	N/A
[25]	S2	✓	N/A	N/A	N/A	N/A
[43]	S3	✓	✓	✓	N/A	N/A
[31]	S4	N/A	N/A	N/A	N/A	Pearson correlation coefficient
[10]	S5	✓	✓	✓	✓	N/A
[32]	S6	N/A	✓	✓	✓	N/A
[44]	S7	✓	✓	✓	✓	N/A
[13]	S8	✓	✓	✓	✓	N/A
[33]	S9	✓	✓	✓	✓	N/A
[45]	S10	✓	✓	✓	✓	N/A
[14]	S11	✓	N/A	N/A	N/A	N/A
[27]	S12	✓	N/A	N/A	✓	N/A
[15]	S13	✓	N/A	N/A	N/A	N/A
[38]	S14	✓	N/A	N/A	N/A	N/A
[39]	S15	✓	N/A	N/A	✓	N/A
[16]	S16	✓	N/A	N/A	N/A	N/A
[34]	S17	N/A	N/A	N/A	✓	N/A
[17]	S18	✓	✓	✓	✓	N/A
[28]	S19	✓	N/A	N/A	✓	N/A
[18]	S20	✓	✓	✓	✓	N/A
[29]	S21	N/A	N/A	N/A	✓	N/A
[46]	S22	✓	N/A	N/A	N/A	N/A
[47]	S23	N/A	N/A	N/A	✓	N/A
[35]	S24	✓	N/A	N/A	N/A	N/A
[36]	S25	N/A	✓	N/A	✓	N/A
[19]	S26	✓	N/A	N/A	N/A	N/A
[30]	S27	N/A	✓	✓	✓	N/A
[20]	S28	✓	✓	✓	✓	Roc-area
[26]	S29	N/A	N/A	N/A	✓	N/A
[37]	S30	✓	N/A	N/A	✓	N/A
[21]	S31	✓	✓	✓	✓	N/A
[22]	S32	N/A	✓	✓	N/A	N/A
[48]	S33	✓	N/A	N/A	✓	N/A
[49]	S34	✓	✓	✓	✓	N/A
[50]	S35	✓	✓	✓	✓	N/A
[23]	S36	✓	✓	✓	N/A	N/A
[24]	S37	✓	✓	✓	✓	N/A
[3]	S38	✓	✓	N/A	✓	N/A

Table 6 Research results reported from the collected research papers

Ref.	Study	Results				
		Accuracy	Recall	Precision	F-measure	Other
[12]	S1	93.25%	91.92%	94.48%	N/A	N/A
[25]	S2	79.13%	N/A	N/A	N/A	N/A
[43]	S3	Lexicon:66.50%	Lexicon:84.50%	Lexicon:48.50%	N/A	N/A
		ML: 70%	ML: 35%	ML: 47%		
[31]	S4	N/A	N/A	N/A	N/A	PCC: 69.20%
[10]	S5	96%	90.2%	92.8%	91.5%	N/A
[32]	S6	N/A	86%	94%	83%	N/A
[44]	S7	84.16%	84.03%	84.84%	84.15%	N/A
[13]	S8	88.72%	93.11%	92.10%	88.27%	N/A
[33]	S9	96%	94%	95%	95%	N/A
[45]	S10	95.81%	91.15%	90.22%	90.68%	N/A
[14]	S11	75.2%	N/A	N/A	N/A	N/A
[27]	S12	95%	N/A	N/A	93%	N/A
[15]	S13	85.03%	N/A	N/A	N/A	N/A
[38]	S14	94.33%	N/A	N/A	N/A	N/A
[39]	S15	95.4%	N/A	N/A	93.4%	N/A
[16]	S16	75.9%	N/A	N/A	N/A	N/A
[34]	S17	N/A	N/A	N/A	72.14%	N/A
[17]	S18	57.8%	57%	70%	63%	N/A
[28]	S19	85%	N/A	N/A	84.95%	N/A
[18]	S20	99.96%	99.96%	99.96%	99.96%	N/A
[29]	S21	N/A	N/A	N/A	96%	N/A
[46]	S22	95%	N/A	N/A	N/A	N/A
[47]	S23	N/A	N/A	N/A	88%	N/A
[35]	S24	93.5%	N/A	N/A	N/A	N/A
[36]	S25	N/A	78.5%	N/A	75.90%	N/A
[19]	S26	93.5%	N/A	N/A	N/A	N/A
[30]	S27	N/A	99.7%	90.5%	87.6%	N/A
[20]	S28	82.1%	84%	85%	84%	87%
[26]	S29	N/A	N/A	N/A	69.9%	N/A
[37]	S30	65.05%	N/A	N/A	64.46%	N/A
[21]	S31	88%	98.34%	100%	88.08%	N/A
[22]	S32	N/A	98%	80%	N/A	N/A
[48]	S33	88.2%	N/A	N/A	87.8%	N/A
[49]	S34	82.7%	81%	82.2%	81.2%	N/A
[50]	S35	94%	93.8%	93.9%	93.9%	N/A
[23]	S36	95.6%	95.07%	96.02%	N/A	N/A
[24]	S37	85.28%	80%	75%	63.95%	N/A
[3]	S38	93%	93%	N/A	92%	N/A

Note that only the best results are included in this table

Table 7 Platforms used to construct a dataset of Arabic dialects

Ref.	Study	Resources platform	Ref.	Study	Resources platform
[12]	S1	Dataset: JEERAN website	[18]	S20	Twitter
[25]	S2	Facebook	[29]	S21	Movie reviews from: elcinema.com and cairo360.com
[43]	S3	Twitter	[46]	S22	Twitter
[31]	S4	Twitter	[47]	S23	Twitter
[10]	S5	Twitter	[35]	S24	Twitter, book review website and other Social Media sites
[32]	S6	Facebook and web pages	[36]	S25	Twitter
[44]	S7	Twitter	[19]	S26	Twitter
[13]	S8	Twitter	[30]	S27	Twitter
[33]	S9	Twitter	[20]	S28	Twitter
[45]	S10	Apple Store and Google Play	[26]	S29	Twitter
[14]	S11	Twitter and book reviews	[37]	S30	Twitter
[27]	S12	Twitter, Facebook, public survey and mystery shopper	[21]	S31	Twitter
[15]	S13	Twitter	[22]	S32	Facebook
[38]	S14	Twitter, book reviews, movies reviews, and blogs	[48]	S33	Social Media websites
[39]	S15	Hotel's reviews	[49]	S34	Facebook and Twitter
[16]	S16	Twitter	[50]	S35	Facebook and Twitter
[34]	S17	Twitter	[23]	S36	JEERAN website
[17]	S18	Arabic book reviews from goodreads.com	[24]	S37	Twitter
[28]	S19	Twitter	[3]	S38	Twitter

tweets, and Twitter is a popular social media platform in the Arab world. However, it is essential to investigate other popular social media platforms for constructing resources for Arabic Dialect texts.

5 Conclusion

Sentiment analysis in texts is one of the most important applications of the NLP field of science. Different approaches have been used so far for Sentiment Analysis of Arabic Texts. In this systematic review, the author explored the researches and

studies that were in the field of Sentiment Analysis of Arabic Dialects. The author reported that: in most of the papers authors didn't specify the dialect type, and mostly it was mentioned (Arabic Dialects), 50% of the utilized datasets in Sentiment Analysis of Arabic Dialects are constructed by the authors themselves and their sizes are between 10,000 and 50,000, machine learning algorithms (classifiers) are the most common approach that was used, the best result for Sentiment Analysis of Arabic Dialects achieved while different machine learning algorithms (classifiers) were used, and Twitter is the most commonly used social media platform for constructing datasets for Arabic Dialects texts.

Acknowledgements This work is a part of a project submitted to The British University in Dubai.

References

1. Bayari R, Bensefia A (2021) Text mining techniques for cyberbullying detection: state of the art. Adv Sci Technol Eng Syst J 6(1):783–790
2. Alshamsi A, Bayari R, Salloum S (2020) Sentiment analysis in English texts. Adv Sci Technol Eng Syst 5(6):1638–1689. https://doi.org/10.25046/AJ0506200
3. Abu I, Magdy W (2021) A comparative study of effective approaches for arabic sentiment analysis. Inf Process Manag 58(2):102438
4. Al Shamsi AA, Abdallah S (2021) Text mining techniques for sentiment analysis of arabic dialects: literature review text mining techniques for sentiment analysis of arabic dialects: literature review. Adv Sci Technol Eng Syst J 6(1):1012–1023. https://doi.org/10.25046/aj0601112
5. Abuata B, Al-Omari A (2015) A rule-based stemmer for Arabic Gulf dialect. J King Saud Univ Comput Inf Sci 27(2):104–112
6. Kilmer PD (2010) Review article: review article. J Pract Criticism 11(3):369–373. https://doi.org/10.1177/1461444810365020
7. Larabi Marie-Sainte S, Alalyani N, Alotaibi S, Ghouzali S, Abunadi I (2019) Arabic natural language processing and machine learning-based systems. IEEE Access 7:7011–7020
8. Khrisat AA (2015) Arabic dialects and classical Arabic language Adv Soc Sci Res J 2(3):254–260
9. Soliman AB, Eissa K, El-Beltagy SR (2017) AraVec: a set of arabic word embedding models for use in Arabic NLP. Procedia Comput Sci 117:256–265
10. Mustafa HH, Mohamed A, Elzanfaly DS (2017) An enhanced approach for arabic sentiment analysis. Int J Artif Intell Appl 8(5):1–14
11. Khalifa S, Habash N, Abdulrahim D, Hassan S (2016) A large scale corpus of Gulf Arabic. In: Proceedings of the 10th international conference on language resources and evaluation, LREC 2016, pp 4282–4289
12. Al-Harbi O (2019) A comparative study of feature selection methods for dialectal Arabic sentiment classification using support vector machine, vol 19, issue 1, pp 167–176
13. Alomari KM, Elsherif HM, Shaalan K (2017) Arabic tweets sentimental analysis using machine learning. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 10350, pp 602–610
14. Abu Kwaik K, Dobnik S, Chatzikyriakidis S (2019) Can modern standard arabic approaches be used for arabic dialects? Sentiment analysis as a case study. In: The 3rd workshop on arabic corpus linguistics, pp 1–11

15. El-Beltagy SR, Khalil T, Halaby A, Hammad M (2018) Combining lexical features and a supervised learning approach for Arabic sentiment analysis. In: Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). vol 9624, pp 307–319. https://doi.org/10.1007/978-3-319-75487-1_24
16. Rizkallah S, Atiya A, ElDin Mahgoub H, Heragy M (2018) Dialect versus MSA sentiment analysis. Adv Intell Syst Comput 723:605–613
17. Al-Ayyoub M, Nuseir A, Kanaan G, Al-Shalabi R (2016) Hierarchical classifiers for multi-way sentiment analysis of arabic reviews. Int J Adv Comput Sci Appl 7(2):531–539. https://doi.org/10.14569/ijacsa.2016.070269
18. Gamal D, Alfonse M, El-Horbaty ESM, Salem ABM (2018) Implementation of machine learning algorithms in Arabic sentiment analysis using N-gram features. Procedia Comput Sci 154:332–340. https://doi.org/10.1016/j.procs.2019.06.048
19. Alfonse M, Salem A (2015) Opinion mining for Arabic dialects on Twitter, 2018, September
20. Atoum JO, Nouman M (2019) Sentiment analysis of Arabic Jordanian dialect tweets. Int J Adv Comput Sci Appl 10(2):256–262
21. Ibrahim MA, Salim N (2013) Sentiment analysis of Arabic tweets: with special reference restaurant tweets. Int J Comput Sci Trends Technol 4(3):173–179
22. Mdhaffar S, Bougares F, Estève Y, Hadrich-Belguith L, Mdhaffar S, Bougares F, Estève Y, Sentiment LH (2018) Sentiment analysis of Tunisian dialects: linguistic resources and experiments to cite this version: HAL Id: hal-01592418
23. Al-Harbi O (2017) Using objective words in the reviews to improve the colloquial arabic sentiment analysis, vol 6, issue 3. https://doi.org/10.5121/ijnlc.2017.6301
24. Al-Azani S, El-Alfy ESM (2017) Using word embedding and ensemble learning for highly imbalanced data sentiment analysis in short Arabic text. Procedia Comput Sci 109:359–366. https://doi.org/10.1016/j.procs.2017.05.365
25. Mataoui M, Zelmati O, Boumechache M (2016) A proposed lexicon-based sentiment analysis approach for the vernacular Algerian Arabic. Res Comput Sci 55–70
26. Al-Twairesh N, Al-Khalifa H, Alsalman A, Al-Ohali Y (2018) Sentiment analysis of Arabic tweets: feature engineering and a hybrid approach
27. Hamdi A, Shaban K, Zainal A (2018) CLASENTI: a class-specific sentiment analysis framework. ACM Trans Asian Low Resour Lang Inf Process 17(4)
28. Elshakankery K, Ahmed MF (2019) HILATSA: a hybrid Incremental learning approach for Arabic tweets sentiment analysis. Egypt Inf J 20(3):163–171
29. Brahimi B, Touahria M, Tari A (2019) Improving sentiment analysis in Arabic: a combined approach. J King Saud Univ Comput Inf Sci
30. Duwairi RM (2015) Sentiment analysis for dialectical Arabic. In: 2015 6th international conference on information and communication systems, ICICS 2015, pp 166–170
31. AlZoubi O, Tawalbeh SK, AL-Smadi M (2020) Affect detection from Arabic tweets using ensemble and deep learning techniques. J King Saud Univ Comput Inf Sci. https://doi.org/10.1016/j.jksuci.2020.09.013
32. Guellil I, Mendoza M, Azouaou F (2020) Arabic dialect sentiment analysis with ZERO effort. Case study: Algerian dialect. Intel Artif 23(65):124–135
33. Oussous A, Benjelloun FZ, Lahcen AA, Belfkih S (2020) ASA: a framework for Arabic sentiment analysis. J Inf Sci 46(4):544–559. https://doi.org/10.1177/0165551519849516
34. Gridach M, Haddad H, Mulki H (2018) Empirical evaluation of word representations on Arabic sentiment analysis. Commun Comput Inf Sci 782(October):147–158. https://doi.org/10.1007/978-3-319-73500-9_11
35. Abu Kwaik K, Saad M, Chatzikyriakidis S, Dobnik S (2019) LSTM-CNN deep learning model for sentiment analysis of dialectal Arabic. Commun Comput Inf Sci 1108:108–121. https://doi.org/10.1007/978-3-030-32959-4_8
36. Alali M, Mohd Sharef N, Azmi Murad MA, Hamdan H, Husin NA (2019) Narrow convolutional neural network for Arabic dialects polarity classification. IEEE Access 7:96272–96283

37. Heikal M, Torki M, El-Makky N (2018) Sentiment analysis of Arabic tweets using deep learning. Procedia Comput Sci 142:114–122. https://doi.org/10.1016/j.procs.2018.10.466
38. Omara E, Mosa M, Ismail N (2019) Deep convolutional network for Arabic sentiment analysis. In: Proceedings of the Japan-Africa conference on electronics, communications, and computations, JAC-ECC 2018, (January), pp 155–159. https://doi.org/10.1109/JEC-ECC.2018.8679558
39. Al-Smadi M, Qawasmeh O, Al-Ayyoub M, Jararweh Y, Gupta B (2018) Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. J Comput Sci 27:386–393. https://doi.org/10.1016/j.jocs.2017.11.006
40. Alsayat A, Elmitwally N (2019) A comprehensive study for Arabic sentiment analysis (challenges and applications). Egypt Inf J 4–9
41. Mallek F, Belainine B, Sadat F (2017) Arabic social media analysis and translation. Procedia Comput Sci 117:298–303. https://doi.org/10.1016/j.procs.2017.10.121
42. Almuqren L, Cristea AI (2016) Framework for sentiment analysis of Arabic text. In: HT 2016—proceedings of the 27th ACM conference on hypertext and social media, pp 315–317
43. El-Masri M, Altrabsheh N, Mansour H, Ramsay A (2017) A web-based tool for Arabic sentiment analysis. Procedia Comput Sci 117:38–45. https://doi.org/10.1016/j.procs.2017.10.092
44. Aldayel HK, Azmi AM (2015) Arabic tweets sentiment analysis—a hybrid scheme. J Inf Sci 42(6):782–797. https://doi.org/10.1177/0165551515610513
45. Areed S, Alqaryouti O, Siyam B, Shaalan K (2020) Aspect-based sentiment analysis for Arabic government reviews. Stud Comput Intell 874:143–162
46. Alayba AM, Palade V, England M, Iqbal R (2018) Improving sentiment analysis in Arabic using word representation. In: 2nd IEEE international workshop on Arabic and derived script analysis and recognition, ASAR 2018, pp 13–18
47. Mustafa Qamar A, Alassaf M (2020) Improving sentiment analysis of Arabic tweets by one-way ANOVA. J King Saud Univ Comput Inf Sci
48. Mulki H, Haddad H, Gridach M, Babaoglu I (2020) Syntax-ignorant N-gram embeddings for dialectal Arabic sentiment analysis. Nat Lang Eng
49. Mulki H, Haddad H, Gridach M, Babaoğlu I (2019) Syntax-ignorant N-gram embeddings for sentiment analysis of Arabic dialects, pp 30–39. https://doi.org/10.18653/v1/w19-4604
50. Mulki H, Haddad H, Ali CB, Babaoglu I (2018) Tunisian dialect sentiment analysis: a natural language processing-based approach. Computacion y Sistemas 22(4):1223–1232

Crowdsourced NLP Retraining Engine in Chatbots

Mohammad Arif Nezami and Rakhshanda Rukham

Abstract Chatbots became a common customer service and e-commerce tool in recent years. Most of the chat platforms offer chatbot integration opportunities. Natural Language Processing (NLP) are powering chatbots worldwide and are at a well-developed stage for English and major languages. Still, lots of improvement and regular retraining required as use cases for chatbots can be very different and evolving every day. Retraining is also important for languages that don't have very strong NLP. In this paper, a crowdsourced, user-generated retraining system is being suggested to make the chatbot NLPs auto-retrained by the users who are actually using it. Experiments result in an average of 17% improvement in the NLP results.

Keywords Chatbots · NLP · Training · Retraining · AI · Crowdsource

1 Introduction

Chatbots are the most common Artificial Intelligence-based product that has been adopted by the mass people. Chatbots use Natural Language Process which is a technique to understand human language by machines [1]. Based on previous learnings, models first computes a relevance score and based on the score an action is taken [2]. Famous scientist Alan Turing is known for extensively working on machine intelligence. He started with the questions "Can machines think?" and then tried to define "machine" and "think" to get to a conclusion. Later, the process became popular as the "Turing Test" [3]. Making machines talk with human beings started there and are still going on. Almost all the popular chat platforms including

M. A. Nezami (✉)
United International University, Dhaka, Bangladesh
e-mail: mnezami201075@mscse.uiu.ac.bd

R. Rukham
Preneur Lab Youth and Innovation Trust, Dhaka, Bangladesh
e-mail: rakhshanda@preneurlab.com

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_26

Facebook Messenger, WhatsApp, Viber etc. provides developers with the opportunity to connect chatbots via APIs and webhooks. Chatbots are one of the most successful products of machine learning techniques. Businesses and organizations around the world use chatbots to communicate with customers. This has become the most effective low-cost customer support platform as well. By 2025, the chatbot market will be at least 3 Billion US Dollars globally. This huge market opens up opportunities for research and development in the area [4].

There is a lot of research happening on the topic for the last many years now. A chatbot is a mature topic [5]. Still, there are always areas of new innovation and improvements. Chatbot retraining is one of those. All NLP engines depend on the training dataset to make sure it is working as expected [6]. As chatbots are now universal and every day new words, new short forms of existing words, emojis, and phrases are coming out [7], it is important to keep the chatbot NLP engine updated with a new set of data every day. Doing it manually is time-consuming and costly. The goal of this technique is to make sure chatbots can be intelligent enough to learn from user interaction itself. There are a lot of NLP methods that are being widely used to create chatbots. Mostly, all the NLP based chatbots are trained first with a set of data and tested before launching it for its users. There will be always a gap between the test dataset and what users will do during a real conversation. It is required to keep retraining the system to make sure the bot is able to answer queries relevant to the users. The bot can be trained manually. But takes a lot of work hours to identify and train the cases where the bot failed to answer.

We have described a process to auto-retrain with the help of users using the chatbot. The process is simple and also improves the user experience and lessens the robotic monotonous tone of the chatbot. Results describe an increase in customer experience scores for chatbots using the process.

2 Related Works

Crowdsourcing techniques replaced many traditional ways of doing things in a number of areas [8]. Researchers in [9], describes the possibilities of crowdsource in machine learning. This mentions working examples of crowdsourcing techniques used in the real world. The paper discusses four (4) major applications but doesn't include chatbot neither provides any specific models or results. According to [10], crowdsourcing can also be used for taxonomy creation in different use cases.

Deep learning techniques mentioned in [11] that can train chatbots unsupervised. The book [1] provides a thorough journey of enterprise-greade chatbots. This also pictures an architect of a task oriented chatbot as well as provides descriptions of different machine learning algorithms to build a better chatbot. This book doesn't cover much on the retraining process after the bot is launched.

There are similar resources available on training and building machine learning chatbots but there are very few resources that talk about post development improvement and re-training which this paper aims at.

3 Methodology

The first step of building an NLP based chatbot is to train the NLP engine that will power the chatbot. Datasets are built based on the use case of the chatbot. The more data is in the dataset the more the NLP engine will be able to answer correctly. Chatbots are trained before launching. As human language and interaction can have numerous forms, chatbots can become robotic and monotonous which decreases user satisfaction. Regular re-training is also done in some systems. These re-trainings require extra resources and time-consuming. Chatbots can become resource-heavy for maintenance and improvement. As users are regularly interacting with the bots, it will be an efficient way if the users' inputs can be fed to the machine to retrain it. Crowdsourced re-training eliminates the difficulties of maintaining a chatbot system after launching.

3.1 Training the NLP Engine

We are using python-based NLTK to understand the process. NLTK is the abbreviation of Natural Language Toolkit. NLTK provides easy to use interfaces to work with human language data. The developer and research community around NLTK makes it rich in resources.

Training process:

1. Load python modules
2. Load dataset
3. Apply Preprocessing in the dataset using NLTK. Such as
    a. Text Cleaning (remove emoji, punctuations, irrelevant symbol)
    b. Tokenization
    c. Stemming
    d. Remove Stopwords
4. Split the preprocessed dataset into a train dataset and test dataset for the deep learning model
5. Create LSTM or Fully connected Neural Network (Densely connected) deep learning model
6. Fit train: A data to the deep learning
7. Train the model
8. Validate the model with test data
9. Saved the required data as the pickle file and the trained model for further test

3.2 Testing Process

Testing is undoubtedly an important part to make sure the NLP engine is ready to go in the reproduction or not. The steps of the testing process are as below.

1. Load required python module
2. Load pickle data and saved model
3. Apply preprocessing step in the dataset using nltk
4. Fit unseen text at the loaded model for intent classification
5. Output result: intent with its probability score

3.3 Building the Web App

As we know python will require to be set up as a Web app to make sure it can work with external data. Flask and Django are two popular frameworks for building python based web applications. We are using the Flask framework in this project.

Python applications will stop automatically when you command line. You have to make sure you use nohop command to keep the server up and running all the time.

3.4 Connecting with Webhook

With the NLP we need to add some extra lines of code to make sure we have one webhook that we can connect with Facebook or any chat platform we are working with.

The simplest workflow of a chatbot normally looks like the flow showed at Fig. 1.

Bot replies are normally done with CURL replies to platform-specific APIs with access tokens and security measures in the format allowed by the platforms.

3.5 Deep Dive in the NLP Output

Normally the NLP engine will reply to the possible match with the user query with the intent and entities it was trained with. It generally provides a score which is between 0 and 1. 0 means the least possible match and 1 means the highest possible match.

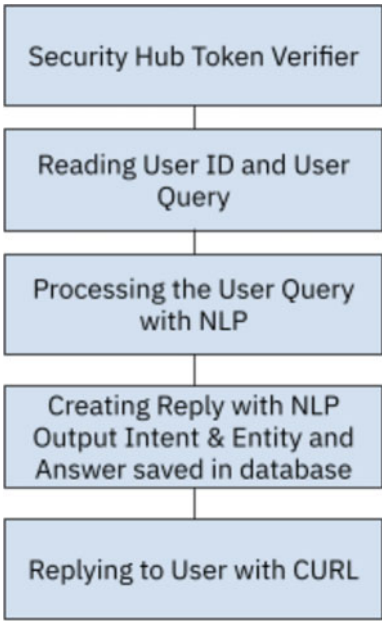


Fig. 1 Simple chatbot workflow

If the score is more than 0.9 or even more than 0.8 it is a very good match. The bot should reply considering this. If the match is below 0.8 (Fig. 2) then it is confusing and there should be some mechanism to tackle that.

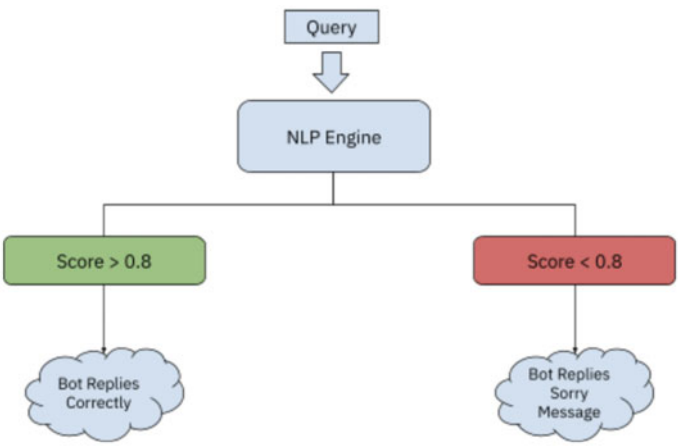


Fig. 2 Normal chatbot decision making process

4 Proposed Solution

We are proposing a solution that will utilize crowdsourced user-generated replies to retrain the model (Fig. 3). This will help the bot learn automatically with user interactions with any need for admin or bot owner interactions.

4.1 (Step 1) Showing 3 Possible Replies

The NLP Engine will provide the score to whatever extent it can match with the intents. It is being suggested to show users 3 options to choose from the top 3 scorer intents. For example, the user typed "What is the lowest price but beautiful shoe available in your store?" The NLP engine output is as shown in Table 1.

The proposed model is showing all the 3 options to the user and asking what the user prefers to know (Fig. 4).

This will help the user as well because the guesses are close enough. So at this point, the user can easily choose one.

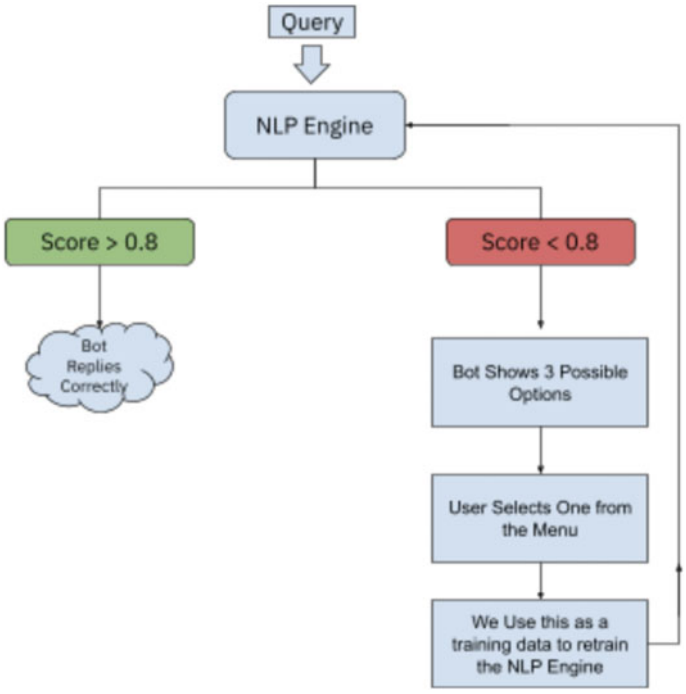


Fig. 3 Proposed decision making process

Table 1 Confidence score of keywords

SL	Intent	Score
1	Low price	0.80
2	Beautiful shoe	0.78
3	Available stock	0.70

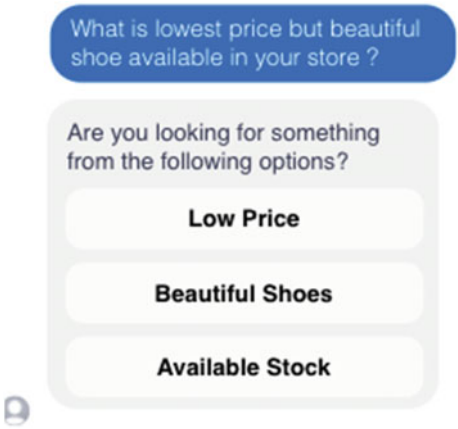


Fig. 4 Screenshot of chatbot showing 3 suggestions

4.2 (Step 2) Crowdsourced Learning Process

When the user chose one of the three (3) close options (Fig. 5), we will use that to retrain the bot. Even if the user doesn't choose from those we can use the learning that can be something else.

4.3 Retraining the Bot: Phase 1

The process shown at the last step will continue to happen whenever the bot is not sure about the results. If the bot is very popular and being used by many users, it will have a good amount of user-generated retraining data regularly.

1. Load required python modules
2. Load the existing dataset
3. Load all the new user texts and intents
4. Append user text and intent to existing dataset-A and name it as dataset-A (update)
5. After successfully data append, use dataset-A (update) for retraining process
6. So the after data insertion final dataset is dataset-A (update)

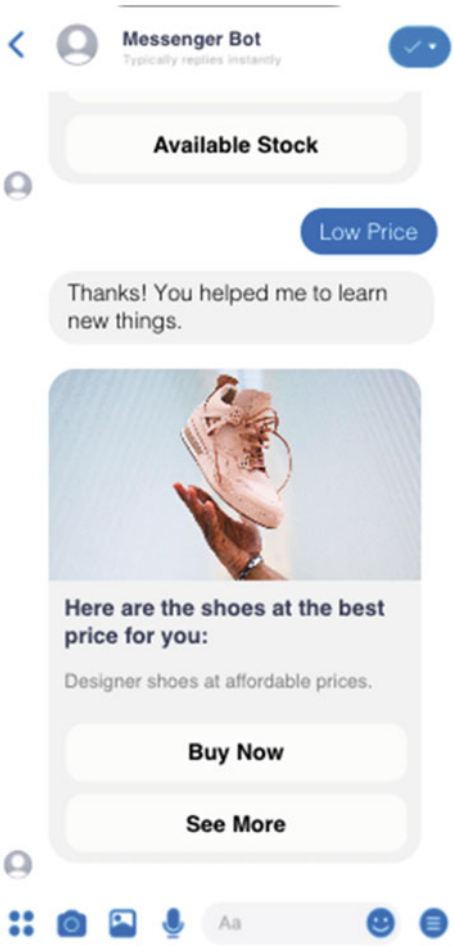


Fig. 5 Screenshot of user selecting one option from suggested options

4.4 Retraining the Bot: Phase 2

The update in the training dataset will be a regular process. The learning will grow whenever the bot fails to answer any query. But the retraining is suggested to be held at one specific time every day. This is to make sure the update is regular but also that shouldn't hamper the process. A cronjob can be set for a time when there is less or almost no user interaction with the chatbot. Preferably early morning around 3 am.

1. Load python modules
2. Load pre-insertion dataset-A (update) as a new dataset
3. Apply Preprocessing in the dataset using nltk. Such as

a. Text Cleaning (remove emoji, punctuations, irrelevant symbol)
b. Tokenization
c. Stemming
d. Remove Stopwords

4. Split the preprocessed dataset into a training dataset and test dataset for the deep learning model
5. Create an LSTM or Fully connected Neural Network (Densely connected) deep learning model
6. Fit train: A data to the deep learning
7. Train the model
8. Validate the model with test data
9. Saved required data as a pickle file and the trained model for further test

5 Results and Discussion

This technique showed extraordinary performance in 90 days of testing in production chatbots. Three (3) production chatbots of three (3) different industries were tested with the process mentioned in the paper. Users' feedback was fed to the machine for retraining. No manual re-training was done for the chatbots. On average, the bot performance increased by 17% without any external human agent training. The results can be realized from Table 2.

This process not only improves future performance but also provides a better experience for the users as they can select from available options of the bot in comparison to keep typing new sentences. This improves ICE (Instant Customer Experience) Score on average by 25% compared to chatbots where the process is not being followed. ICE scores are based on user satisfaction after a chat session is completed. As having a human agent is costly, this process simplifies the NLP training process especially for the languages that have little NLP resources available.

Table 2 Results from experiments

	Correct reply to users with first dataset	Correct reply to users after user generated retrained dataset after…		
		1 month	2 months	3 months
Chatbot 1: topic financial institution chatbot	70%	85%	89%	91%
Chatbot 2: topic NGO chatbot	78%	87%	89%	92%
Chatbot 3: topic news portal chatbot	75%	81%	85%	91%

6 Conclusion

Chatbots have come a long way. It became one of the most important tools as a linkage between organizations and customers. Chatbots are also the most common artificial intelligent tool available for enterprise till the date. Improvement of the chatbots will help millions of organizations, businesses, brands and users around the world. This will reduce tons of enterprise cost and improve customer satisfaction.

References

1. Galitsky B (2019) Developing enterprise chatbots: learning linguistic structures. Springer. ISBN 3030042995:9783030042998
2. Jurafsky D, Martin JH (1999) Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition, 2nd edn. Prentice Hall Inc Publications
3. Turing AM (1950) Computing machinery and intelligence. Mind LIX(236):433–460. https://doi.org/10.1093/mind/lix.236.433
4. businesswire.com (2020) The world's aI-based chatbot market, 2020–2025—conversational aI will improve telemarketing efficiency. https://xho.to/gwl9doed
5. Nagarhalli TP, Vaze V, Rana NK (2020) A review of current trends in the development of chatbot systems. In: 6th international conference on advanced computing and communication systems (ICACCS)
6. Das S, Kumar E (2018) Determining accuracy of chatbot by applying algorithm design and defined process. In: 4th international conference on computing communication and automation (ICCCA)
7. Samal BR, Behera AK, Panda M (2017) Performance analysis of supervised machine learning techniques for sentiment analysis, 2017. In: Third international conference on sensing, signal processing and security (ICSSS)
8. Lasecki WS, Miller CD, Naim I, Kushalnagar R, Sadilek A, Gildea D, Bigham JP (2017) Scribe: deep integration of human and machine intelligence to caption speech in real-time. Commun ACM 60(11)
9. Abhigna BS, Soni N, Dixit S (2018) Crowdsourcing—a step towards advanced machine learning. Procedia Comput Sci 132:632–642. https://doi.org/10.1016/j.procs.2018.05.062
10. Bragg J, Mausam, Weld DS (2013) Crowdsourcing multi-label classification for taxonomy creation. In: HCOMP
11. Karri SPR, Kumar BS (2020) Deep learning techniques for implementation of chatbots. In: International conference on computer communication and informatics (ICCCI)

Early Detecting Students at Risk Using Machine Learning Predictive Models

Ahlam Wahdan, Sendeyah Hantoobi, Mostafa Al-Emran, and Khaled Shaalan

Abstract Analytics in educational environments has received much attention during the last few years. Maintaining a high retention rate is still a major concern in higher educational institutions. Therefore, this research aims to early detect students at risk using three machine learning predictive models, namely support vector machine (SVM), neural networks (NN), and K nearest-neighbors (KNN), based on a new dataset collected from 800 students through surveys. The criteria used to evaluate the models were accuracy, sensitivity, and specificity. Regarding the accuracy, the SVM model has outperformed the NN and KNN models, where it achieves 86.7%. Concerning the sensitivity, the NN model was more sensitive to detect failure cases than the other two models. Regarding the specificity, it was very high for the three models. It is believed that the results could assist the educators in early detecting students at risk, and therefore, reducing students' dropouts.

Keywords Learning analytics · Students at risk · Predictive models · Machine learning · Neural networks · Support vector machine · K nearest-neighbors

1 Introduction

Nowadays, higher educational institutions are experiencing pressure and complexity due to the increasing competition among them. An acceptable amount of research is conducted with the aim of improving the pedagogical activities and increasing the number of community-based learning environments [1]. In such learning environments, the user usually utilizes online tools, such as discussion

A. Wahdan · S. Hantoobi · M. Al-Emran (✉) · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE
e-mail: ahlam.wahdan@buid.ac.ae

S. Hantoobi
e-mail: 20192120@student.buid.ac.ae

K. Shaalan
e-mail: khaled.shaalan@buid.ac.ae

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_27

forums, online chats, and learning management systems. Using those online activities creates a huge amount of data called "big data", which cannot be processed with the traditional analytical methods [2, 3]. This creates the need for exploring new techniques to process this type of data.

Analyzing the data and finding out the interesting patterns require a system or application [4, 5]. This system/application can read the data in a more comprehensive and meaningful way for the organization by producing useful information that will help in decision-making. This is called "Business Intelligence". One of the main components in "Business Intelligence" is analytics. This component provides techniques to find out the trends and patterns in data that will lead to decision-making. Learning analytics goals can be summarized in improving the learning context and helping learners to succeed. Learning analytics is evolved to explore the big data capabilities in education. Essentially, it collects data from different resources and analyzes them to produce more descriptive information or trends that correlate the attributes with the behavior of student's success. These attributes are used as predictors for student's success and could help in determining students at-risk. Learning analytics systems can send alarms to all stakeholders in case there is a need to intervene in such cases. These alarms can be sent to the student at-risk to inform him/her about the requirements needed to succeed. Also, it can be sent to the educators to notify them about their students at-risk to handle the situation [6].

Increasing the retention rate and the graduation rate is a crucial issue for any university [7]. Early predicting students at-risk provides the educational institutions' stakeholders the chance to intervene [8]. The instructor can redesign his course curriculum and choose the suitable learning aids. Further, the institution administration can take decisions from the beginning of the semester, and the students can be notified early so they would be encouraged to change their behavior. It has been noticed that there is a lack of research in terms of identifying students at-risk more professionally and feasibly. In line with the previous literature [9–11], it is imperative to indicate that machine learning algorithms can predict those students and provide informative decisions. Therefore, this research aims to predict students at-risk at an early stage before the semester starts using three predictive models, namely Support Vector Machine (Gaussian RBF SVM), Neural Network Backpropagation (NN BP), and K Nearest-Neighbors (KNN).

2 Literature Review

Table 1 shows a summary of the relevant studies by depicting the variables that were used in building the predictive models.

Table 1 Summary of relevant studies

Source	Variables	Purpose
[12]	The number of homework problems solved, the performance on homework problems, the results of weekly paper tests, the number of times students log into the learning platforms, the number of days students watch videos, the number of days students participate in online conversations, and the number of times the front and back searches are clicked while watching videos	To predict the students' academic performance by considering their experience in a calculus course that combined online learning and traditional learning
[13]	The number of times users logged into the learning management system, time spent within the system, the number of downloads, discussions, and dialogues between students, the number of exams that were taken, the number of forums students participated in, among other variables	A multicollinearity test is used to identify the variables that measure the same results. The Pearson correlation coefficient is used to measure the relationship between the dependent and independent variables. The regression analysis is employed to determine the effectiveness of the model in predicting students' outcomes
[14]	Course success and feedback engagement	The study used a series of trackable Call to Action (CTA) connections to evaluate the link between course success and feedback engagement in addition to predicting the students' perceptions
[15]	VLE activity—forum usage	Comparing the students' usage of VLE with other students who performed better in the same course
[16]	Grades, major subject, students' certainty of choosing a certain subject, and contribution of parents in the financial issue	A developed dashboard that uses certain variables and then predicts the students who will return the following year to complete their studies. The students who will not return will be identified as at-risk
[17]	Number of assignments, tests submitted, number of VLE course sessions, and number of forums the students read and contributed in	Logistic regression is used to identify students in good standing or at-risk

3 Research Methodology

3.1 Research Design

According to the reviewed literature, we have formulated survey questions to collect the data using survey monkey to build the dataset. Secondly, in the data preprocessing step, we made cleansing for the data from the null values and some

preprocessing for the data, such as normalization. After preprocessing, we have made a descriptive analysis which included finding the correlation between the independent and dependent variables so we can find out the most significant variables for the dependent ones. After that, building predictive models took place using different data mining techniques, including SVM, NN, and KNN. Lastly, tuning the algorithms to increase the accuracy and validating the models to find out the best fit for predicting the course score.

3.2 Dataset

The sample for this study has been collected using a survey where it covered around 800 students. The questions were focused on the academic track of specific courses, university cumulative GPA, school GPA, and demographic questions. The sample is believed to be significant and valuable because it has been collected from multinational students where they have studied their high schools and colleges in different countries, such as Jordan, India, Lebanon, Egypt, etc. Also, it covers information related to the whole track of specific subjects. This means that we have all the prerequisite courses' marks for the course that we want to predict.

3.3 Preprocessing

In the preprocessing step and the cleansing process, we have removed all the records that have a lot of nulls, and we ended up with 400 records. We have then checked out the null values, and we substituted them with the average grade for each variable. In terms of normalization, the data collected varies in scales. The prerequisite courses are on a scale of letter grades (i.e., A, B, C, and D). We have then converted them into numerical variables, given that the commutative GPA is on a scale of 4 points. We have unified the scales of all variables by normalization and making the scale to be varied from 0 to 1 to make the analysis easier.

3.4 Building the Predictive Models

For all the models that we have built, we followed the steps shown in Fig. 1.

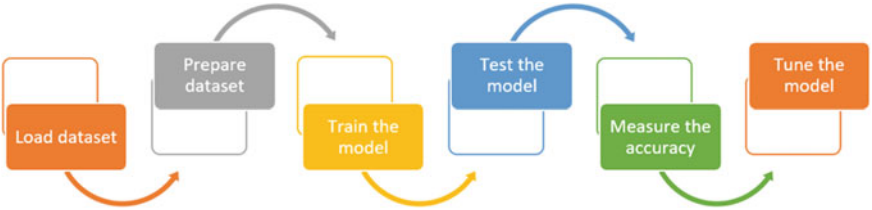


Fig. 1 Steps for building predictive models

Table 2 Confusion matrix

Actual	Predicted	
Attrition	True positive (TP)	False negative (FN)
No attrition	False positive (FP)	True negative (TN)

3.5 Validating the Predictive Models

We have used popular metrics to evaluate the performance of the predictive models. These metrics can be calculated in many ways, but in our research, we have used the confusion matrix. The confusion matrix is used in binary classification to tabulate the correct and incorrect prediction [18], as shown in Table 2. In that, there are four terminologies, including TP, TN, FP, and FN. True Positive (TP) denotes the positive outcome when it is predicted correctly. True Negative (TN) denotes the negative outcome when it is predicted correctly. False Positive (FP) denotes the positive outcome when it is not predicted correctly. False Negative (FN) denotes the negative outcome when it is not predicted correctly. The metrics that we used in our research are accuracy, sensitivity, and specificity, as shown in the following equations [19].

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

4 Results and Discussion

4.1 Descriptive Statistics

Table 3 shows the mean and the standard deviation for the variables that were calculated after the normalization process.

4.2 Correlation Analysis

We have used the Pearson equation to find out the correlation between the independent and dependent variables. If the correlation value is greater than 0.7, it indicates a significant correlation. If the correlation value between 0.5 and 0.7, this indicates a moderate correlation. As per the results in Table 4, the cumulative GPA has a significant correlation with all the courses except Math 1. Regarding the Statistics course and its prerequisite courses, we found that there is a moderate correlation.

Table 3 Descriptive statistics

Course	Mean	Standard deviation
Cumulative GPA	0.743	0.123
Math1	0.776	0.172
Math2	0.776	0.165
Linear algebra	0.770	0.160
Differential equations	0.712	0.178
Statistics	0.746	0.171

Table 4 Correlation coefficient results

	Cumulative GPA	Math1	Math2	Linear algebra	Differential equations	Statistics
Cumulative GPA	1	0.65	0.7	0.7	0.7	0.71
Math1		1	0.6	0.4	0.72	0.56
Math2			1	0.5	0.66	0.56
Linear algebra				1	0.66	0.62
Differential equations					1	0.63
Statistics						1

4.3 Predictive Models

4.3.1 SVM Model

In SVM, there are different methods of the kernel. We have used the SVM RBF to build our model, which is the most common method in SVM regression. It is argued that SVM regression is used in mapping the data into a higher dimension space [20]. The function we used for training is KSVM. In KSVM, two parameters affect the accuracy, including cost and Gamma. It is imperative to mention that the cost of misclassification is C. If we choose a small value of C, the misclassification will be small, and the margin will be soft. On the other hand, choosing a large value of C, the misclassification will be high, and the margin is called a hard margin. In this case, the model will be stricter, and it will cause overfitting.

Regarding Gamma, a small value of Gamma makes the model more linear and constrained. So, if the value is set to be very small, this will lead to difficulty in capturing the shape or the complexity of the data. Therefore, we used the tune function to choose the best parameter values of C and Gamma. The summary function will return the best parameter value. Then, we train our model again using the best parameter values (Cost = 10 and Gamma = 0.01). Accordingly, we test our model on the fold that we did not use in training using the predict function. We have then assessed the model by finding the accuracy, specificity, and sensitivity using the confusion matrix. The accuracy of the SVM model is 86%, but the sensitivity and specificity are 0 and 1, respectively.

4.3.2 NN Model

In NN, there are different methods. We have used the NN Backpropagation to build our model, which is an optimization algorithm that repeats two cycles for the input vector. There are several factors affecting the accuracy of NN, such as learning rate, weight, activation function, number of neurons in the hidden layers, and the number of the hidden layers. To increase the accuracy of NN, we tuned the number of the hidden layers and the number of neurons by increasing them. However, this increases the complexity of the model and increases the training time [20]. After that, we test our model on the fold that we did not use in training using the predict function. We have then assessed the model by finding the confusion matrix. The results showed that the accuracy is 83%, the sensitivity is 31%, but the specificity is 92%.

4.3.3 KNN Model

We have used the trainContol function to make cross-validation for our dataset. We have divided the dataset into three sets. To train the model using the KNN algorithm, we have used the train function and specified method “knn”. We have also

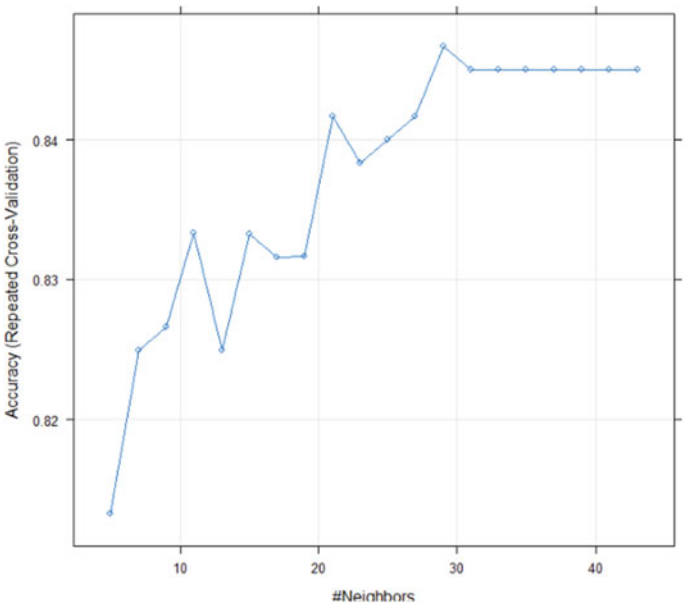


Fig. 2 KNN model accuracy

specified the control parameters using 'trControl'. Figure 2 shows the accuracy, and it is plotted using the plot function. It can be seen that the best accuracy achieved is when the number of neighbors k = 29. Further, we have tested the model on the fold that we did not use in training using the predict function. We have then assessed the model by finding the confusion matrix where the accuracy is 83%, the sensitivity is zero, and the specificity is 98%.

5 Conclusion and Future Work

Recent studies have focused on finding new methods that increase students' achievement and retention rate [21, 22]. Therefore, excellence in teaching has been on the top priority for higher educational institutions. During the last decade, several universities started to use learning management systems that create a digital footprint [23], and a massive amount of data that is called big data. Big data cannot be analyzed using traditional methods like descriptive statistics. So, learning analytics evolved to enable us to look at traditional reporting to give us a new vision in predicting what will happen in the future. Thus, decision-makers can use these predictions to make decisions.

As per the reviewed literature, we found that some studies tried to predict students at risk. However, the approach used in those studies relied on using

student's activities in a specific course so that the prediction will be available only after the third or half of the term. This makes intervening to avoid student's failure very late. To handle this issue, this research is an attempt to provide the educational institution stakeholder the chance to intervene at an early stage. In that, the instructor can redesign his/her course curriculum and use the appropriate learning aids. Further, the institution administration can take decisions from the beginning of the term, and the student can be notified early and encouraged to change his/her behavior. All of these early interventions could enhance the students' retention rate.

In line with the previous arguments, this research aimed to predict the course score and students at risk before the course start. To achieve this aim, we have used three learning analytics techniques, including Support Vector Machine (Gaussian RBF SVM), Neural Network Backpropagation (NN BP), and K Nearest-Neighbors (KNN) using R software. The criteria used to evaluate the models were accuracy, specificity, and sensitivity. Regarding the accuracy, the SVM model has outperformed the NN and KNN models, where it achieves 86.7%. Concerning the sensitivity, the NN model was more sensitive to detect failure cases than the other two models. Regarding the specificity, it was very high for all the models.

It is imperative to report that the sensitivity is zero in KNN and 30 in NN. This can be enhanced by balancing the data, as the dataset was not balanced because the ratio of the failure to pass students in the Statistics course is around 1:4. This can be enhanced in future research by trying to balance the data using SMOTE analysis and then build the model again. Further, more variables could be embedded, such as indicators for students' psychological states like interests and motivation, university rank, and students' specializations.

Acknowledgements This work is a part of students' project submitted to The British University in Dubai.

References

1. Al-Qaysi N, Mohamad-Nordin N, Al-Emran M (2021) Developing an educational framework for using whatsapp based on social constructivism theory. In: Recent advances in intelligent systems and smart applications. Springer, pp 243–252
2. Marks A, Al-Ali M (2016) Analytics within UAE higher education context. In: 2016 3rd MEC international conference big data smart city, ICBDSC 2016, pp 368–373. https://doi.org/10.1109/ICBDSC.2016.7460396
3. Hantoobi S, Wahdan A, Al-Emran M, Shaalan K (2021) A review of learning analytics studies. In: Recent advances in technology acceptance models and theories, pp 119–134
4. Mhamdi C, Al-Emran M, Salloum SA (2018) Text mining and analytics: a case study from news channels posts on Facebook, vol 740
5. Salloum SA, Al-Emran M, Abdallah S, Shaalan K (2017) Analyzing the Arab Gulf newspapers using text mining techniques. In: International conference on advanced intelligent systems and informatics, pp 396–405. https://doi.org/10.1007/978-3-319-64861-3_37
6. Oakleaf M (2016) Getting ready & getting started: academic librarian involvement in institutional learning analytics initiatives. J Acad Librariansh 42(4):472–475. https://doi.org/10.1016/j.acalib.2016.05.013

7. Kondo N, Okubo M, Hatanaka T (2017) Early detection of at-risk students using machine learning based on LMS log data. In: 2017 6th IIAI international congress on advanced applied informatics (IIAI-AAI), Nov 2017, pp 198–201. https://doi.org/10.1109/IIAI-AAI.2017.51
8. Al-Shabandar R, Hussain AJ, Liatsis P, Keight R (2019) Detecting at-risk students with early interventions using machine learning techniques. IEEE Access 7:149464–149478. https://doi.org/10.1109/ACCESS.2019.2943351
9. Al-Emran M, Al-Kabi MN, Marques G (2021) A survey of using machine learning algorithms during the COVID-19 pandemic. In: Emerging technologies during the era of COVID-19 pandemic. Springer Science and Business Media Deutschland GmbH, pp 1–8
10. Arpaci I, Al-Emran M, Al-Sharafi MA, Shaalan K (2020) A novel approach for predicting the adoption of smartwatches using machine learning algorithms. In: Recent advances in intelligent systems and smart applications. Springer, pp 185–195
11. Zaza S, Al-Emran M (2015) Mining and exploration of credit cards data in UAE. In: Proceedings—2015 5th international conference on e-learning, ECONF 2015, pp 275–279. https://doi.org/10.1109/ECONF.2015.57
12. Lu OHT, Huang AYQ, Huang JCH, Lin AJQ, Ogata H, Yang SJH (2018) Applying learning analytics for the early prediction of students' academic performance in blended learning. Educ Technol Soc 21(2):220–232
13. Mwalumbwe I, Mtebe JS (2017) Using learning analytics to predict students' performance in moodle learning management system: a case of Mbeya University of science and technology. Electron J Inf Syst Dev Ctries. https://doi.org/10.1002/j.1681-4835.2017.tb00577.x
14. Iraj H, Fudge A, Faulkner M, Pardo A, Kovanović V (2020) Understanding students' engagement with personalised feedback messages. In: ACM international conference proceeding series, pp 438–447. https://doi.org/10.1145/3375462.3375527
15. JISC (2016) Analysing use of the VLE at the University of Maryland, Baltimore County
16. JISC (2016) Identifying at-risk students at New York Institute of Technology
17. JISC (2016) Transferring predictive models to otherinstitutions from Marist College
18. Arpaci I, Huang S, Al-Emran M, Al-Kabi MN, Peng M (2021) Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. Multimed Tools Appl 80:11943–11957. https://doi.org/10.1007/s11042-020-10340-7
19. Barros TM, Neto PAS, Silva I, Guedes LA (2019) Predictive models for imbalanced data: a school dropout perspective. Educ Sci 9(4). https://doi.org/10.3390/educsci9040275
20. Huang S (2011) Predictive modeling and analysis of student academic performance in an engineering dynamics course, Utah State University
21. Saa AA, Al-Emran M, Shaalan K (2019) Factors affecting students' performance in higher education: a systematic review of predictive data mining techniques. Technol Knowl Learn 24 (4):567–598. https://doi.org/10.1007/s10758-019-09408-7
22. Saa AA, Al-Emran M, Shaalan K (2019) Mining student information system records to predict students' academic performance. In: International conference on advanced machine learning technologies and applications, pp 229–239
23. Al-Nuaimi MN, Al-Emran M (2021) Learning management systems and technology acceptance models: a systematic review. Educ Inf Technol 1–35. https://doi.org/10.1007/s10639-021-10513-3

Chaotic Particle Swarm Optimization Based on Meeting Room Approach for Designing Bijective S-Boxes

Hussam S. Alhadawi, Sinan Q. Salih, and Yasir Dawood Salman

Abstract This paper proposed a modified and enhanced multi-swarm Particle Swarm Optimization algorithm based on Meeting Room Approach (MPSO) with Tent Logistic map. The proposed modified MPSO algorithm was used to generate strong 8×8 substitution boxes (S-Box) while a Chaotic Tent map was used to update the velocities of each leader in the meeting room to improve their movements and enhance the exploration capability of the MPSO. The generated S-box was validated using cryptographic performance evaluation criteria, such as nonlinearity, bijectivity, strict avalanche criteria, bit independence criteria, differential probability, and linear approximation probability. The obtained results showed that the proposed algorithm performed better than some of the state-of-the-art algorithms and proved its capability against cryptanalysis attacks.

Keywords Particle swarm optimization (PSO) · Meeting room approach (MRA) · S-boxes · Chaotic maps · Cryptography

1 Introduction

In the current digital era, the confidentiality of messages is normally protected using block ciphers. The choice of block ciphers in protecting data integrity is driven by their substitution box (S-box) component which, as a nonlinear component, provides the confusion property to the corresponding block cipher as noted by [1]. The confusion concept portrays the importance of concealing the overall framework of

H. S. Alhadawi (✉) · Y. D. Salman
Computers Technologies Engineering Department, Dijlah University College, Baghdad, Iraq
e-mail: hussam.alhadawi@duc.edu.iq

Y. D. Salman
e-mail: yasir.dawood@duc.edu.iq

S. Q. Salih
Computer Science Department, Dijlah University College, Baghdad, Iraq
e-mail: sinan.salih@duc.edu.iq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_28

cryptosystems. Hence, a cryptographically strong S-box is important for achieving a confusion property that can resist linear cryptanalysis and differential attacks [2, 3]. A careful selection of the S-boxes is important to keep such ciphers from easy breakage; therefore, many studies have focused on the design of cryptographically-strong S-boxes [4]. Even though the design of an S-box that can exhibit all cryptographic properties is impossible, it is important that certain trade-offs be established [5].

Metaheuristics have been implemented for handling different types of optimization problems, such as machine learning [6–8], Engineering Problems [9], Data Clustering, and Green Energy [10–13]. Nevertheless, many metaheuristic-based design algorithms (such as firefly algorithm) have been proposed for constructing 8×8 S-boxes [14]. For instance, a discrete chaotic map-based cuckoo search algorithm was proposed by [15]. While the Sine cosine optimization process for S-box design was conducted by [16]. Furthermore, a globalized firefly algorithm for retrieving strong S-box design was developed by [17]. but being that no single metaheuristic-based S-box can be considered the best over the existing ones, studies are still ongoing for novel metaheuristic-based S-box strategies [15]. One of the most popular swarm intelligence metaheuristics is Particle Swarm Optimization (PSO) [18], it has been extensively used for solving optimization problems. Recently, a modified version of the PSO algorithm based on a multi-swarm architecture was proposed for solving numerical and engineering optimization problems [19], the proposed approach was named Multi-Swarm Particle Swarm Optimization (MPSO).

In this paper, the Chaotic Tent map was used to improve the performance of the MPSO algorithm; hence, the proposed system in this work was named "Tent Multi-Swam Particle Swarm Optimization (TMPSO)". The rest of the paper was structured as follows: Sect. 2 explained the proposed TMPSO algorithm, while Sect. 3 presented the obtained results. The conclusion of the study was presented in Sect. 4.

2 The Proposed Algorithm (TMPSO)

Many researchers have shown interest in PSO since its development in 1995 because it is an easy and simple optimization technique [18]. Despite undergoing several modifications, PSO is still suffering from the issues of premature convergence, slow convergence speed, poor quality of the solution, and update strategy. PSO was first developed as a nature-inspired metaheuristic that was inspired by the birds' flocking behavior. In the PSO, the flock of birds is considered randomly distributed within an area with just one food source.

The position of the food source is unknown to the birds in the flock, but it is strategically placed at a known distance from each bird. The nearest bird to the food piece communicates with the other birds in the flock, prompting them to move towards the food source. Therefore, the food source is considered the optimal value

while each bird in the flock is considered a particle whose distance from the food piece represents the value of the objective function, making it possible to model the flocking pattern of the birds as a function optimization process.

The concept of PSO is based on the fact that the velocity and position of each particle in the solution space are specified during an active search for the optimal solution to any NP problem. The updating of the position of the particles is done with respect to the currently achieved local and global optima. Hence, the updating of the position of each particle (e.g., particle *i*) is expressed thus:

$$X_i(t+1) = X_i(t) + V_i(t+1) \tag{1}$$

where t represent the current status, $t + 1$ represent the post-update status, (t + 1) represent the velocity of the new particles. Note: The time difference $\Delta t = (t + 1) - 1$ is a time unit and the velocity of particle i is expressed by:

$$V_i(i+t) = \omega V_i(t) + c_1 r_1 \left(X_i^P - X_i(t)\right) + c_2 r_2 (X_G - X_i(t)) \tag{2}$$

where $v_i(t)$ represent the current velocity of the particle, X_i^P is the particles' local best position while XG is the global best particle position; ω, c_1, and c_2 are constants determine the relevance of each component of the velocity equation, while r_1 and r_2 are random values in the range of [0, 1].

A modified version of PSO based on multi-swarm architecture was proposed recently [20, 21]. The architecture was called the "Meeting Room Approach (MRA)", where several swarms interact and share information to discover better positions in the search space. MRA was built-in another recently proposed algorithm called Nomadic People Optimizer (NPO) [22], where, the leaders of the clans share their positions with the overall best leader. However, the mathematical models were different.

In the MPSO, the best particle for each swarm is selected and sent to the meeting room for the selection of the overall best leader. In the meeting room, the positions of the other leaders are updating using the following mathematical model:

$$W^N = \left(\frac{W^G - W^N}{T}\right) \times Rand(0,1) \tag{3}$$

$$v_i^N(t+1) = W^N \times v_i^N(t) + rc\left(X_i^G - X_i^N(t)\right) \tag{4}$$

$$X_i^N(t+1) = X_i^N(t) + v_i^N(t+1) \tag{5}$$

where W^N and W^G represent the inertia weight for the best and normal leaders inside the meeting room, while X_i^N, X_i^G, and v_i^N represent the position of the normal and best leaders, and the velocity of the normal leader, respectively; r is a random value in the range of [0, 1], and c is a constant that is equal to the social parameter (i.e., c_2).

The above mathematical model showed that the velocity is the core of updating the position of each normal leader towards the best leader; however, the moving steps depend on the random value of r. In the standard version of MPSO, r is generated via uniform distribution, which may generate random values in near ranges and may decrease the chances of the global search. In this study, a Chaotic Tent Map was used instead of the uniform distribution for generating r as follows:

$$R_{i+1} = \begin{cases} \frac{x_i}{0.7} & if\ R_i < 0.7 \\ \frac{10}{3}(1 - R_i) & otherwise \end{cases} \tag{6}$$

where R_i = a real value ranging from 0 to 1 and represents a single dimension of any given problem.

The evaluation/fitness function used in this study for evaluating each particle in all swarms is the standard nonlinearity for each obtained S-box, which is calculated as follows:

$$\mathrm{N}_f = 2^{n-1}\left(1 - 2^{-n} \max_{w \in GF(2^n)} \left|WS_{\langle f \rangle}(w)\right|\right) \tag{7}$$

where the computation of the Wash spectrum of $f_{(x)}$ is done thus:

$$WS\langle f \rangle = \sum_{x \in GF(2^n)} (-1)^{f(x) \oplus x.w}$$

where $w \in GF(2^n)$ and $x.w$ are the dot product of x and w that is provided as follows:

$$error = 112 - NL(f_i) \tag{8}$$

where 112 represents the optimal nonlinearity value in Advanced Encryption Standard (AES) [23]. The nonlinearity function of the particle agent is denoted by NL. The best agent fitness is selected based on their lowest error value for the current iteration.

2.1 Adjustment

The detection of the 8 × 8 S-box in the considered objective function can be done using a 256 items sequence. This will ensure that the generated S-box achieved bijectivity, meaning that each S-box item must be different. Meanwhile, fulfilling this condition is not always possible for the particle generated from swarm operations; hence, the following adjustments must be made to the S-box of each particle:

1. Each S-box item must be scanned completely and if during the scanning process, it is observed that any value is repeated in the S-box, the position of the value that was repeated in the V sequence must be saved. Excluding the first item, any other empty item will substitute all other items with similar values. This scan can also aid in determining the absence of any S-box item.
2. Any S-box item that is missing during a random search will be sorted and their positions must be saved; the sequence following the sorted items must be defined as V2.
3. For the V sequence, the position is randomly filled from sequence V2.

3 Performance Analysis of the Proposed S-Box

For the proposed algorithm, the parameter setting was as follows: number of swarms = 5, size of the swarm = 10, number of iterations = 100, $w = 0.74$, $c_1, c_2 = 1.42$. Table 1 showed the generated S-box in this study. The analysis of the security of the S-box generated in this study was based on various state-of-the-art evaluation metrics identified by security experts for the validation of the strength of new S-boxes. The study by [2], presented the differential uniformity concept through its differential cryptanalysis for DES to coin the notion of differential probability computed from I/O XOR distribution. Linear cryptanalysis was also performed by [3] to reflect the importance of linear approximation probability.

Table 1 S-box generated by MPSO algorithm

#	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	102	88	65	69	156	97	77	122	95	121	134	131	117	90	71	164
1	22	176	106	136	58	112	206	221	28	37	132	133	103	56	159	163
2	49	79	87	251	187	46	63	233	19	94	247	162	167	41	109	192
3	104	138	165	98	129	186	123	85	128	181	153	205	139	34	29	36
4	150	9	74	170	55	140	50	119	222	78	24	126	62	120	5	157
5	145	100	127	111	6	188	242	174	114	169	207	155	68	189	130	180
6	246	118	47	66	107	250	223	110	219	80	196	60	61	220	14	31
7	249	105	241	168	253	215	76	148	243	202	212	171	194	45	144	82
8	191	226	113	27	108	42	23	152	161	83	149	197	135	43	143	158
9	204	10	236	35	200	32	26	52	7	67	137	217	8	146	30	184
A	147	193	17	151	227	54	231	15	254	166	238	198	0	40	230	53
B	13	18	25	81	237	232	179	213	125	252	141	44	218	175	51	225
C	178	173	203	75	115	21	248	234	160	208	182	172	1	48	116	33
D	240	86	91	244	210	64	185	211	12	154	101	92	99	224	39	209
E	93	183	73	38	195	3	89	142	59	57	199	216	72	201	124	4
F	177	16	235	70	96	84	2	228	229	20	245	190	214	255	239	11

Dawson and Tavares [24] applied the information theory in the S-box design criteria. Hence, the strength of S-boxes is commonly assessed based on 6 major metrics which are the bijective property, strict avalanche criteria (SAC), nonlinearity, maximum expected linear probability (MELP), and bit independence criteria (BIC) differential uniformity. The performance of the generated S-box in this study was compared to that of a few S-boxes generated with metaheuristic-based methods [14, 25–27].

3.1 Bijective Property

This method was first introduced by [28] to confirm bijective property. The S-box is said to satisfy the bijectivity property when the Boolean function f_i are required to be balanced, meaning that its outputs should have equal distribution of 0's and 1's. A Boolean function is determined with $f_i(1 \leq i \leq n)$ as:

$$wt\left(\sum_{i=1}^{n} a_i f_i\right) = 2^{n-1} \tag{9}$$

where $a_i \in \{0, 1\}$, $(a_1, a_2, \ldots, a_n) \neq (0, 0, \ldots, 0)$, and $wt()$ represents the Hamming weight. The bijectivity of 8×8 is satisfied if its look-up contains different values within interval $[0, 255]$. The generated S-boxes based on the proposed method satisfied the bijective property.

3.2 Nonlinearity

Nonlinearity is important for achieving plaintext confusion and for better immunity of block ciphers against linear attacks. For any Boolean function, the nonlinearity measure determines the possible least distance from all the Boolean functions; this is practically calculated from the Walsh spectrum as presented in Eq. 7. Table 2 presents the nonlinearity score of the S-box generated in this study. Observably, all the nonlinearity values were greater than 106, suggesting that the generated S-box can withstand linear cryptanalysis.

Table 2 Nonlinearity score of the proposed S-box

S-box	N_1	N_2	N_3	N_4	N_5	N_6	N_7	N_8	Min	Avg
NL	108	108	108	108	108	108	108	110	108	108.25

Table 3 SAC of the proposed S-box

0.4688	0.4844	0.4844	0.4531	0.5000	0.5000	0.4844	0.4844
0.4219	0.4531	0.4688	0.4688	0.5781	0.5313	0.4844	0.5156
0.5000	0.5156	0.4531	0.4844	0.4844	0.5156	0.4531	0.4688
0.4688	0.5469	0.4844	0.4688	0.5000	0.4531	0.4844	0.5313
0.4375	0.5156	0.5156	0.5156	0.4531	0.5156	0.5000	0.4531
0.4688	0.5000	0.4688	0.4531	0.4844	0.5625	0.4375	0.5469
0.5781	0.5000	0.5000	0.5469	0.5313	0.5000	0.4844	0.4688
0.4375	0.4531	0.4531	0.5156	0.5000	0.5156	0.5625	0.4844

3.3 SAC

The first definition of strict avalanche criteria was given by Webster [29]; it was defined as a 50% chance of a change in output bits upon complementation of a single input bit. SAC is calculated using the dependence matrix as given in [30]; the aim of this method is to verify the satisfaction of the 0.5 value by all elements in the S-box. The SAC value of the generated S-box in this work is presented in Table 3, with the average value being 0.491463 which is close to recommended 0.5 value.

3.4 BIC

This was proposed by [29]. for the stipulation of the pair-wise independence of all sets of generated vectors series through single plaintext complementation. The correlation between items is calculated to determine the pair-wise independence between such pairs. Assume an S-box whose Boolean function is $h_1, h_2, \ldots, h_n$; an S-box that meets the BIC criterion, $h_j \oplus h_k (j \neq k, 1 \leq j, k \leq n)$, has been earlier stated to meet the SAC and is expected to be highly nonlinear [31]. Tables 4 and 5 presented the average BIC-SAC and BIC-nonlinearity values respectively, for the generated S-box in this study.

Table 4 BIC-nonlinearity of the proposed S-box

–	102	102	106	104	106	104	106
102	–	104	108	94	102	102	106
102	104	–	106	104	106	102	104
106	108	106	–	100	102	100	108
104	94	104	100	–	104	106	104
106	102	106	102	104	–	100	106
104	102	102	100	106	100	–	100
106	106	104	108	104	106	100	–

Table 5 BIC-SAC of the proposed S-box

–	0.5134	0.5156	0.5067	0.529	0.5335	0.5089	0.4844
0.4955	–	0.5067	0.5000	0.5045	0.4888	0.5022	0.5134
0.5223	0.5022	–	0.5112	0.5156	0.4955	0.5201	0.5134
0.5402	0.5201	0.5156	–	0.5156	0.5045	0.5201	0.5179
0.5000	0.4955	0.4933	0.4866	–	0.4955	0.5089	0.5223
0.5112	0.5223	0.5223	0.4754	0.5379	–	0.4688	0.4844
0.5000	0.4978	0.5313	0.5156	0.4866	0.5112	–	0.5089
0.5246	0.5045	0.4866	0.4821	0.5045	0.4688	0.4665	–

3.5 DP

The nonlinear pair-wise component of S-boxes can be determined by calculating an encryption process. The S-box can exhibit varying uniformity, but a uniform mapping can only be ensured if the input uniquely maps differentially to an output. With these features, the uniform mapping probability for every input bit i is assured. The differential approximation probability technique is used to measure the differential uniformity of an S-box [32]; the process is expressed thus:

$$DP(\Delta x \rightarrow \Delta y) = \left[\frac{\#\{x \in Xs(x) \oplus S(x \oplus \Delta x) = \Delta y}{2^n}\right] \tag{10}$$

where Δx and Δy represents the input/output differentials. The maximum obtained DP value was 10. Which is another evidence that the generated S-box can withstand differential cryptanalysis.

3.6 MELP

For any event, the maximum imbalance value (MIV) can only be measured using the maximal expected linear probability (MELP) which is useful in determining the MIV of an event value. Γx and Γy represent the two masks that are applied to the input and output bits, respectively [33], For a given s-box, the MELP of its bias is given as:

$$LP = \max_{\Gamma x, \Gamma y \neq 0} \left| \frac{\#\{x | x.\Gamma x = s(x).\Gamma y}{2^n} \right| - 0.5 \tag{11}$$

where X contains all the possible inputs while 2^n represent the number of elements in X. Considering the previous metrics, the main goal of S-box designs is to achieve bijectivity and maximize the minimal value for nonlinearity.

Table 6 DP value of the proposed S-box

8	6	6	6	6	6	8	8	6	6	8	6	8	10	8	6
6	8	6	10	8	6	6	8	6	6	6	6	6	6	8	6
6	6	8	8	6	8	6	4	8	6	8	6	6	8	10	6
8	6	6	6	6	6	6	6	6	6	6	6	10	6	6	6
6	8	8	8	6	6	8	6	4	6	6	8	8	8	8	6
6	6	10	6	6	6	6	8	6	8	8	6	8	6	6	8
6	6	6	6	6	6	6	8	8	6	8	6	6	8	8	6
6	8	6	6	6	6	10	8	10	6	6	8	8	6	6	6
6	6	10	6	6	8	6	6	6	6	6	8	6	10	6	6
8	6	6	6	6	6	6	6	6	6	6	6	8	8	8	6
8	8	6	6	8	8	6	6	6	6	6	8	6	6	6	8
8	8	6	6	8	8	6	6	6	6	6	6	8	4	6	8
8	8	6	6	6	6	6	10	6	6	8	6	6	8	6	8
8	8	6	6	6	6	6	6	6	6	6	8	6	6	6	8
8	6	6	6	6	8	8	8	6	8	8	8	8	6	6	8
8	6	6	8	8	6	6	6	6	6	8	6	6	6	8	–

4 Comparison Analysis

In the past years, different optimization algorithm-based approaches have been proposed for designing S-boxes. The S-box construction schemes available in the literature are based on chaos and optimization [14, 25–27]. Table 7 presents a comparison between the proposed method and the available S-boxes in the literature. Several findings were observed from Table 7 and are summarized as follows:

- It is worth mentioning that the proposed S-box obtained the highest and the best average nonlinearity value compared to the other S-boxes. Therefore, the proposed method can generate cryptographically strong S-boxes that can withstand linear cryptanalysis.
- In terms of the SAC test, all the compared S-boxes exhibited good SAC values, meaning that all the obtained values from the S-boxes in the comparison were close to the ideal value of 0.5.
- The proposed s-box achieved a comparable BIC-SAC value with the other comparative S-boxes in this study. However, the BIC-nonlinearity value of the proposed s-box was lower than that of S-boxes in [14, 26], but better than the other S-boxes in the comparison.
- The proposed S-box achieved a maximum DP value of 10 which equaled the value achieved by the comparative S-boxes in Table 7.
- Based on the LP results presented in Table 7, the proposed S-box obtained the same result as the S-box in [14], but better than the other comparative S-boxes. This result is evidence that the proposed method can generate S-boxes that can withstand linear cryptanalysis.

Table 7 The S-box evaluation comparison

S-box	Nonlinearity			SAC	BIC-NL	BIC-SAC	DP	LP
	Min.	Max.	Avg.	Avg.				
Proposed	108	110	108.25	0.4914	103.5	0.5058	10	0.1250
Method in [25]	102	106	104	0.4961	103.28	0.4969	10	0.1406
Method in [26]	106	110	107	0.5015	104.21	0.5016	10	0.1484
Method in [27]	106	110	107.5	0.5093	103.07	0.5029	10	0.1094
Method in [14]	106	108	107.5	0.4943	104.35	0.4982	10	0.1250

5 Conclusion

This paper presented the enhancement of the multi-swarm version of particle swarm optimization (MPSO) algorithm with Chaotic Tent Map for the generation of dynamic S-Boxes. The proposed method in this study was named "Tent Multi-Swarm Particle Swarm Optimization (TMPSO)". The S-Box generated with the proposed algorithm in this work was evaluated using several statistical methods. The experimental results showed that the proposed TMPSO generated a good S-Box with good nonlinearity, SAC, BIC, DP, and LP values. Therefore, the generated S-Box has high cryptographic properties and can withstand different cryptanalysis attacks.

References

1. Shannon CE (1949) Communication theory of secrecy systems. Bell Labs Tech J 28(4):656–715
2. Biham E, Shamir A (1991) Differential cryptanalysis of DES-like cryptosystems. J CRYPTOLOGY 4(1):3–72
3. Matsui M (1993) Linear cryptanalysis method for DES cipher. In: Workshop on the theory and application of of cryptographic techniques, pp 386–397
4. Picek S, Cupic M, Rotim L (2016) A new cost function for evolution of s-boxes. Evol Comput 24(4):695–718
5. Carlet C (2005) On highly nonlinear S-boxes and their inability to thwart DPA attacks. In: INDOCRYPT. Springer
6. Salih SQ (2019) A new training method based on black hole algorithm for convolutional neural network. J Southwest Jiaotong Univ 54(3)
7. Abba S et al (2020) Evolutionary computational intelligence algorithm coupled with self-tuning predictive model for water quality index determination. J Hydrol 124974
8. Afan HA et al (2020) Input attributes optimization using the feasibility of genetic nature inspired algorithm: application of river flow forecasting. Sci Rep 10(1):1–15
9. Chen H et al (2019) A balanced whale optimization algorithm for constrained engineering design problems. Appl Math Model 71:45–59
10. Tao H et al (2020) A newly developed integrative bio-inspired artificial intelligence model for wind speed prediction. IEEE Access 8:83347–83358

11. Hai T et al (2020) Global solar radiation estimation and climatic variability analysis using extreme learning machine based predictive model. IEEE Access 8:12026–12042
12. Mohammed AQ, Al-Anbarri KA, Hannun RM (2020) Introducing newly developed Nomadic People Optimizer (NPO) algorithm to find optimal sizing of a hybrid renewable energy. In: IOP conference series: materials science and engineering. IOP Publishing
13. Mohammed AQ, Al-Anbarri KA, Hannun RM (2020) Optimal combination and sizing of a stand–alone hybrid energy system using a nomadic people optimizer. IEEE Access 8:200518–200540
14. Ahmed HA, Zolkipli MF, Ahmad M (2018) A novel efficient substitution-box design based on firefly algorithm and discrete chaotic map. Neural Comput Appl
15. Alhadawi HS et al (2020) A novel method of S-box design based on discrete chaotic maps and cuckoo search algorithm. Multimedia Tools Appl 1–18
16. Alzaidi AA et al (2018) Sine-cosine optimization-based bijective substitution-boxes construction using enhanced dynamics of chaotic map. Complexity 2018
17. Alhadawi HS et al (2020) Globalized firefly algorithm and chaos for designing substitution box. J Inform Secur Appl 55:102671
18. Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: MHS'95, proceedings of the sixth international symposium on micro machine and human science. IEEE
19. Salih SQ, Alsewari A (2018) Solving large-scale problems using multi-swarm particle swarm approach. Int J Eng Technol 7(3):1725–1729
20. Salih SQ et al (2018) Novel multi-swarm approach for balancing exploration and exploitation in particle swarm optimization. In: International conference of reliable information and communication technology. Springer
21. Salih SQ, Alsewari AA, Yaseen ZM (2019) Pressure vessel design simulation: implementing of multi-swarm particle swarm optimization. In: Proceedings of the 2019 8th international conference on software and computer applications
22. Salih SQ, Alsewari AA (2020) A new algorithm for normal and large-scale optimization problems: nomadic people optimizer. Neural Comput Appl 32(14):10359–10386
23. Daemen J, Rijmen V (1999) AES proposal: Rijndael
24. Dawson M, Tavares SE (1991) An expanded set of S-box design criteria based on information theory and its relation to differential-like attacks. In: Workshop on the theory and application of of cryptographic techniques. Springer
25. Chen G (2008) A novel heuristic method for obtaining S-boxes. Chaos Solitons Fractals 36 (4):1028–1036
26. Ahmad M, Bhatia D, Hassan Y (2015) A novel ant colony optimization based scheme for substitution box design. Procedia Comput Sci 57:572–580
27. Tian Y, Lu Z (2017) Chaotic S-box: intertwining logistic map and bacterial foraging optimization. Math Probl Eng 2017
28. Adams C, Tavares S (1989) Good S-boxes are easy to find. In: Conference on the theory and application of cryptology. Springer
29. Webster A, Tavares SE (1985) On the design of S-boxes. In: Conference on the theory and application of cryptographic techniques. Springer
30. Pieprzyk J, Finkelstein G (1988) Towards effective nonlinear cryptosystem design. IEE Proc E-Comput Digital Tech 135(6):325–335
31. Cui L, Cao Y (2007) A new S-box structure named affine-power-affine. Int J Innovative Comput Inf Control 3(3):751–759
32. Li Y, Wang M (2014) Constructing differentially 4-uniform permutations over GF (2 2m) from quadratic APN permutations over GF (2 2m+ 1). Des Codes Crypt 72(2):249–264
33. Matsui M (1993) Linear cryptanalysis method for DES cipher. In: Workshop on the theory and application of of cryptographic techniques. Springer

Information Security and Networking

Performance of Effect for XPM and FWM in Fiber Optics

Mohammed Salah Abood, Mustafa Maad Hamdi, Bakr Ahmed Taha, Ahmed Shamil Mustafa, and Emad M. Alssaedi

Abstract Optical fiber technologies have revolutionized telecommunication. The need for extensive transmission capability has spurred the production of fiber optics. This study aims to include a definition of XPM and FWM and how they influence the optical signal transmitted through nonlinear dispersive fibers. Consequently, Nonlinear fiber optics plays a significant role in implementing high-capacity optical networks. In this work, we evaluate the effect of XPM and FWM in fiber optics based on OptiSystem simulation results in this paper. We found the Nonlinear optics deals with the behavior of laser, Nonlinear effects in the optical fiber occurring due to nonlinear refractive index are Cross-Phase Modulation (XPM) and Four-Wave Mixing (FWM). In summary, Nonlinear effects are essential for inventing lasers, amplifiers, switching, multiplexing devices, and demultiplexers.

Keywords OptiSystem · Simulation · XPM · FWM · Fiber optics

M. S. Abood (✉)
School of Information and Electronics Engineering, Beijing Institute of Technology, Beijing, China

M. M. Hamdi · A. S. Mustafa
Department of Computer Engineering Techniques, Al-Maarif University College, Al-Anbar, Iraq

B. A. Taha
UKM—Department of Electrical, Electronic and Systems Engineering, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia
e-mail: p103537@siswa.ukm.edu.my

E. M. Alssaedi
Department of Computer Sciences, University of Technology, Baghdad, Iraq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_29

1 Introduction

Chromatic dispersion, fiber-free nonlinearity, and noise from optic or cross-connections in the field of optical communication networks can rapidly degrade signal quality before signals arrive at their destination. The binary signal logic '0' and '1' is difficult to define, giving rise to transmission errors. Corrupted signals must be substituted for the precision of the original signal to reduce the bit rate. In general, signal recovery strategies can be dividing into three categories: reamplification, restoration, and retiming (3R) [1]. Various optical 3R regenerating methods have been published, e.g., using Semiconductor Optical Amplifier (SOA) using cross-grain or cross-phase modulation [2]. Electromagnet modulation, the internet's exponential expansion, contributes to the quantum leap in the optical fiber network's delivery capabilities. They were due to improved information packing performance of a single pull (more than 2 bits/symbol) than traditional binary modulation (1 bit/symbol). The approach to significantly increasing the effective use of spectral bandwidth in existing optic amplifiers and transmission lines is multi-level modulation [3].

Large cross-phase (XPM) modulation is a cross-phase modulus analysis model (XPM) developed in the first-order perturbation principal framework, based on the dispersion-managed coherent fiber optic systems. For random pulse forms, the XPM variance determined analytically. The number of Gaussian times shift pulses is ideal for the non-Gaussian pulse and gives a robust non-Gaussian pulse approximation and even an explicit XPM variance derivation. The XPM variation is calculated analytically in line with the numerical simulations. So, in specific quantum information systems, XPM is essential for two single-photon pulses. Electromagnetically Induced Transparency (EIT) has achieved significant nonlinearity with minimal losses when XPM is applied. The N-type four-level system is a simple, proven EIT-based XPM scheme [4]. There has also been a demonstration of an updated XPM scheme based on the N-type system [5]. XPM is one of the most noticeable nonlinear effects when several SCM signal carrying wavelengths co-propagate in a dispersive fiber. Moving power from one channel to another creates crosstalk that is counterproductive to authentic communication and should have retained the lowest possible degree in the wavelength domain [6]. So, to reach higher performance, the enormous traffic growth needs continued improvement in optical transmission systems. Increasing capability while retaining the same accessibility requires specialized strategies to eliminate linear and nonlinear impairment of transmission. Compensation of the linear impairments such as CDs and PMD is affordable with optical equalization. Digital equalization CDs are cheap (e.g., fractionally spaced finite impulse response filter). However, it is also difficult to reduce nonlinear Kerr-induced impairs (i.e., self-phase modulation (SPM), XPM, and FWM). Current nonlinear equalizers need a wide range of optical parameters and trigger heavy calculations. Moreover, quasi-static transitions to more dynamic optical networks make transmission systems more complex [7].

FWM to phase conjugation (PC) and plasma diagnosis has applied. The methodology was consisting of a nonlinear reaction to an applied optical field by bonded electrons in the material at longer wavelengths (infrared 10 μm to radio wave 10 m) and showed great potential as a plasma diagnostic [8]. FWM can cause significant errors in multi-wavelength transmission systems using dispersion-shifted fibre (DSF) in the zero-dispersion wavelength (λ0) region. Thus, this phenomenon leads to system loss, although frequency generation use [9]. The effectiveness of this nonlinear interaction depends significantly on the correspondence of phases. A significant parameter for the optical fiber phase misalignment is chromatic dispersion. Different dispersion-shifting and non-dispersion-shifted fibres, where a non-zero chromatic dispersion level, one ps/km/nm in the dispersion-shifted fibre, being treated, were investigated as a feature of FWM efficacy [10]. The challenges lie to deliver a clear idea of the impact that cross-phase modulation (XPM) and four-wave mixing (FWM) facilities have on the propagation of optical signals in a nonlinear dispersion fiber using OptiSystem.

This study aims to include a definition of XPM and FWM and how they influence the optical signal transmitted through nonlinear dispersive fibers. On the other hand, we examine various factors to increase the results of XPM and FWM. The capabilities of OptiSystem can also be improved by easily interacting with a wide range of devices. However, the results showed that the XPM and FWM effects' dispersion in a lossless fiber would reduce. To be successful, it is essential to use a higher bit rate because it results in a higher time value.

2 Literature Review

This section describes works of literature research of XPM and FWM effects in an optical fiber system. Different types of studies have been doing for addressing and optimizing the required Bit rates. However, few methods of analysis have been developing to increase the bit rate and wavelength. The change of wavelengths with all-optic modulation has studied in a fiber powered by two pump waves. The operational mechanism is examined with a modulating all-optic step with two parallel/cross-polar pomp waves to produce a precise up-/down-mounted replica of the optical signal. The conversion efficiency was measure with a 300-m high-linear fiber [11, 12]. Another research [13, 14] shows effective XPM modulation based on a closed-loop double-digit framework. By changing phases of the optical fields, the properties of the dual medium have controlled. This phase based XPM system provides significant phase modulations at low light intensities without the need for laser beam cavities or closes focus. The system detects a phase level transition with two pulses consisting of eight photons, all in cool ferrocene. This research [15] proposed an easy solution with Nonlinearity Splits and a reciprocal recovery step algorithm for reducing the cross-phase effect (XPM). The method used designed to reduce the effects of the wavelength (WDM) system. Also, in [16, 17], a comprehensive four-wave mixing (FWM) study and its required matching phase

conditions are reported in graded-index multimode optical fibres. When fiber length is on the order of a meter or less, and the signal and idler produce at broad frequency distances from a pump, the study is mainly essential for spontaneously frequency conversion and photon pair generation through FWM. An empirical expression has obtained to measure the phase misalignment that results from the distribution of the online waveguide between pump, signal, and idler beams.

Thulium (Tm)-doped laser with all fibre wavelengths; an experimental demonstration has made of four-wave mixing (FWM) of strongly Germanic highly nonlinear (HG-HNLF). With the advantage of a high nonlinearity of the HG-HNLF, the FWM-based intensity-dependent gain is integrated into the laser cavity to reduce the competitiveness of gain in Tm-doped fibre. The room temperature is at a range of 0,86 nm in wavelength due to 50-m HG-HNLF, 9, 22, and 36 lasing lines about 10-dB, 20-dB, and 30-dB [18]. Finally [19, 20], these studies aim to develop birefringent fibres (BF) dipole and combo optical solitons along with the four-wave mixing effect (4 WM). Two kinds of raw mediums, Kerr law and parabolic law, are used. The Choudhury approach method has extended to obtain dark soliton solutions in the light (dipole) optical pulse with sparkling backgrounds. Li's technique used to achieve soliton combo solutions that deliver solitary light waves and dark solitary waves.

3 Methodology

This section describes the methodology of XPM and FWM effects in optical fibre communication based on Opti-System simulation tools. We can examine multiple factors responsible for increasing XPM and FWM results, enables nearly any form of optical communication to be built and different optical network evaluations. It also offers a library of active and passive modules to install and monitor fibre optic communication networks' outcomes. As shown in Fig. 1, a method for XPM analysis have built. The optical Gaussian Pulse Generator's operating wavelength is 1551 nm, and its capacity is different from 20 mW. The fibres are 100 km long. The fiber has an effect of self-phase modulation. The resulting XPM input pulse spectrum attributable to the fiber output is displayed using an Optical Spectrum Analyzer.

The two-channel WDM system shown in Fig. 2 is considered for FWM research. At 1551 and 1549 nm wavelengths, two Continual-Wave Lasers (CW) produce the required light signal. The two-channel WDM multiplexer multiplexes two signals. The fiber is supplied with a multiplexed signal. The machine is initially operated by 0 dBm. The productive fiber area is 64 m^2, and its range is 100 km. The fiber end output of the FWM method is seen by the Optical Spectrum Analyzer.

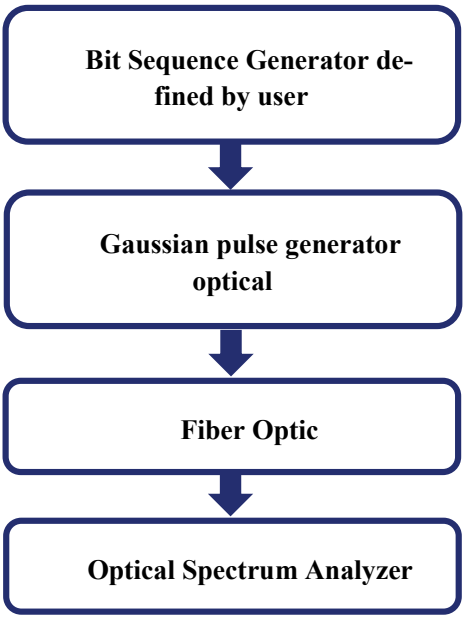


Fig. 1 A simulation model for SPM analysis

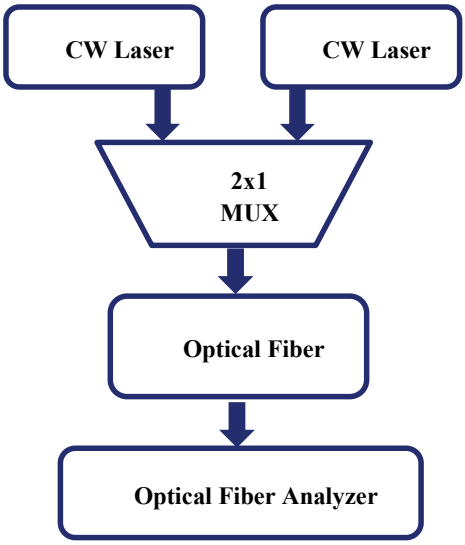


Fig. 2 A simulation model for FWM analysis

4 Simulation Tools

This section describes the simulation of XPM and FWM effects in optical fibre communication based on Opti-System simulation tools. To begin, the layout was created and designed, as shown in Fig. 3. You can see the Numeric tab of the Total Field Non-linear Dispersive Fiber component shown in Fig. 4. The output signal here is many orders of magnitude below the soliton regime since the step size used to simulate soliton propagation is so much smaller. However, it still occupies the (BW) occupied by the signal is more significant and a small enough step size is needed to accurately calculate the 4-wave mixing products resulting from the interaction of the input signals. The step size is determined in this case, rather than

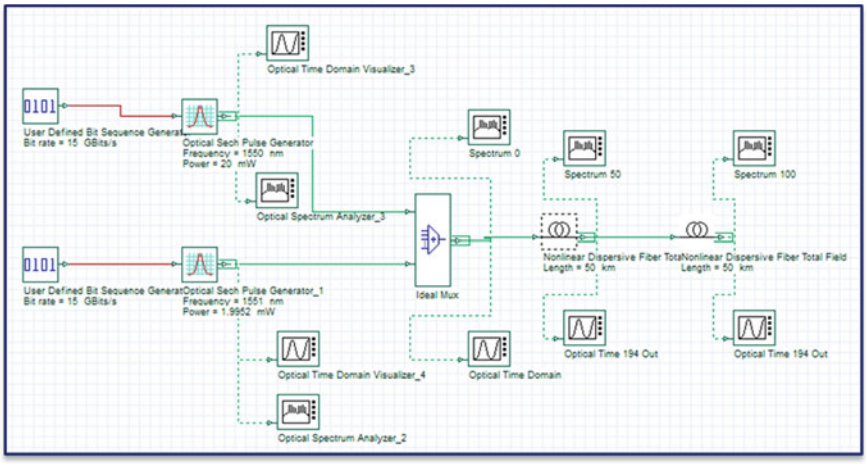


Fig. 3 Create and design layout

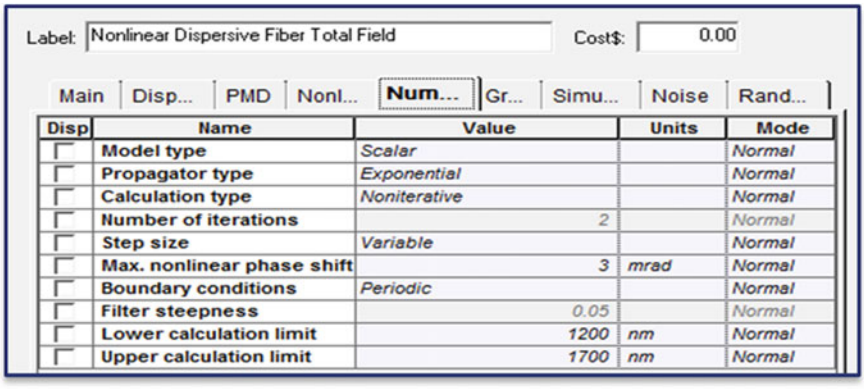


Fig. 4 Nonlinear dispersive fiber total field properties dialogue box-numeric tab

nonlinear, by the coherence length of the FWM process (associated with the self-phase modulation process).

OptiSystem is a comprehensive software architecture suite allowing users in the current optical network transmission layer to design, validate and model visual communications. It is, however, a device-level simulator that focuses on realistic fiber-optic networking simulation. An advanced graphical user interface manages the design of the optical component, network list, models of modules and graphics (GUI). Furthermore, in the physical layer, it simplifies almost any kind of optical connectivity and analyzes a wide variety of lengthy, metropolitan and local networks (LANs). OptiSystem supplies a wide library of prototype optical (OSD) design or simulation and planning data. The OptiSystem capabilities can also be improved by incorporating user elements that interact with a wide range of instruments easily. Any apps like this are available [21] The architecture of the physical layer optical communication networks to the part level.

- CATV or TDM/WDM/CDM network design.
- Passive optical networks (PON) based FTTx.
- Free space optic (FSO) systems and Radio over fiber (ROF) systems.
- SONET/SDH ring design.
- Sender, channel, amplifier, and configuration of the receiver.
- Scatter map concept.
- BER assessment and device punishments for multiple receiver models.
- Expanded BER structure and budget relation calculations.

5 Results

This section describes the results and discussion of XPM and FWM effects in optical fibre communication based on Opti-System simulation tools. The input consists of two Gaussian pulses (800 ps) separated in time and (1 nm) in frequency, as in Fig. 5.

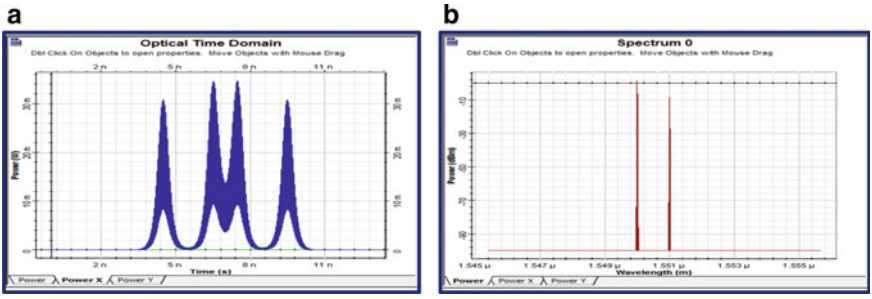


Fig. 5 **a** Input pulses and their spectrum (time), **b** input pulses and their spectrum (wavelength)

The fiber dispersion is = 16 (ps/nm)/km.

The simulation (BW) is about three times greater or higher than the input signal. This is achieved to avoid the FWM products from being aliased when the two pulses communicate non-linearly.

The pulse at (1551 nm) has a maximum of 2 mW power, which means that the pulse at 100 km is insignificant for SPM. However, the pulse (1550 nm) contains 20 mW peak intensity, and for this pulse, the impact of SPM is significant. The XPM would be too big at (1551 nm) pulse. However, the two pulses converge after (50 km) transmission, and the pulse spectra are extended to 1551 nm and back to display Fig. 5. The influence of XPM is responsible for this extension. The spectral expansion of the pulse (1550 nm) is more interesting. The increase is due to SPM for this pulse. The expansion of the pulse (1551 nm) (Cause of XPM), however, because of the presence of GVD, is less. The edge of the stronger pulse moves through, the smaller pulse decreases the gap times and increases the spectral expansion. In (50 km), products from four waves will certainly be mixed at (1549 and 1551 nm). So, the BW signal is the input signal three times. Where both the signal and false frequencies are not appropriate for simulating BW, the FWM papers have been converted to the BW.

At 100 km, when pulses have ceased to overlap and FWM is absent. Even the XPM caused the pulse spectral enhancement (1551 nm). This is attributed to the symptoms associated with pulse increase. The overlap is opposite to pulse separation-related. The interfering pulse with declining amplitude contributes to a frequency transition from the opposite sign to the previous signal from the increasing edge. This paper demonstrates that the dispersion in a lossless fiber will decrease XPM and FWM effects. As pulses begin to overlap in a loss fibre, there is no further symmetry that allows the false waves to disappear, then when they start splitting, and waves that are fake don't disappear. Observed influences and changes in fiber using graphs pulses and their spectrum. The 50 km as in Fig. 6 and 100 km as in Fig. 7 lengths used for our experiment was post-propagation (Fig. 8).

Finally, in this flowchart below, I reflect Microsoft's time and bitrate values, as Fig. 9 shows. We can also see that when the bit rate is higher, the time value is higher.

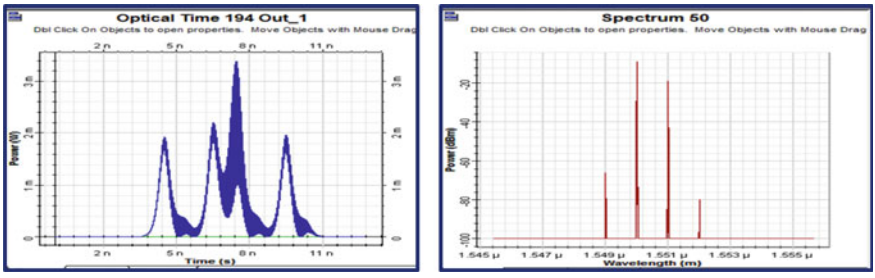


Fig. 6 Pulses and their spectrum in a lossless fiber 50 km after propagation

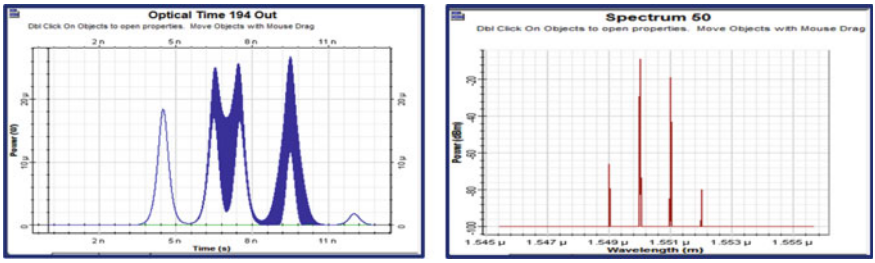


Fig. 7 After 100 km of propagation pulses and spectrum

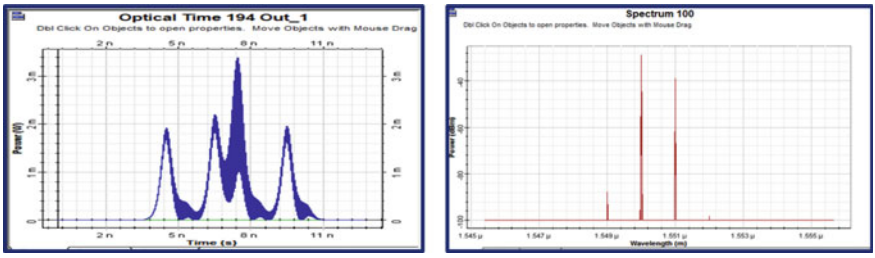


Fig. 8 Including fiber loss

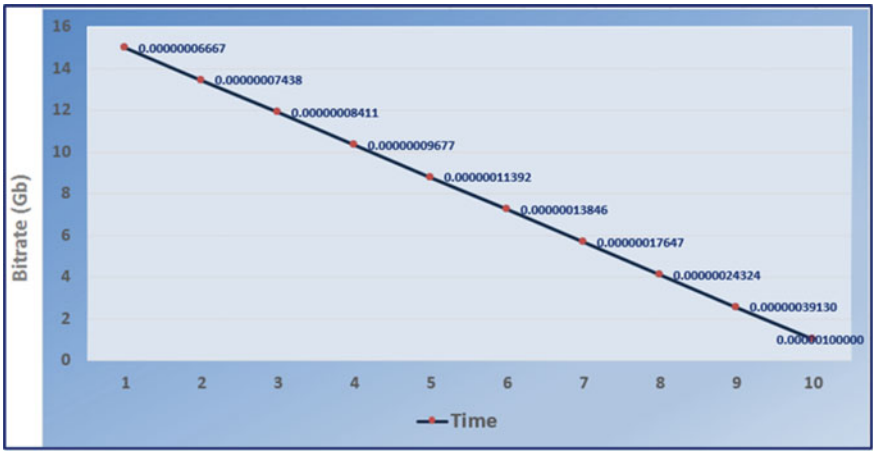


Fig. 9 Time and bitrate

6 Conclusion

Nonlinear effects require a significant interest when designing optical fiber communication systems. Given the above, we conclude that increasing the input power leads to increase XPM and FWM impact at a constant channel spacing. Simultaneously, fibre influences and changes were monitored via graphs, pulse measurements, and the spectrum. The paragraphs showed noticeable results and changes in the fiber. For our experiment, we used 50 and 100 km lengths after post-propagation. OptiSystem is a useful tool to examine the balancing of nonlinear effects. Furthermore, techniques have been presenting for the simultaneous broadband conversion of all-optical simultaneous wavelength by taking advantage of the four-wave mixing (FWM), combined with simultaneous measurement of the nonlinear and Chromatic dispersion component. Finally, nonlinear effects can be useful in solitons generation, pulse compression, amplifiers, lasers, wavelength converters, multiplexing devices, and demultiplexers.

References

1. Hsieh H-S, Feng K-M, Lee M-CM (2010) Study of cross-phase modulation and free-carrier dispersion in silicon photonic wires for Mamyshev signal regenerators. Opt Express 18(9): 9613–9621
2. Leclerc O et al (2003) All-optical signal regeneration: from first principles to a 40 Gbit/s system demonstration. C R Phys 4(1):163–173
3. Leclerc O et al (2003) Optical regeneration at 40 Gb/s and beyond. J Lightwave Technol 21(11):2779
4. Shiau B-W et al (2011) Low-light-level cross-phase modulation with double slow light pulses. Phys Rev Lett 106(19):193006
5. Catanese C et al (2019) A survey of neural network applications in fiber nonlinearity mitigation. In: 2019 21st international conference on transparent optical networks (ICTON). IEEE
6. Mohammed MJ, Hamdi MM (2018) Hybrid tow feature extraction descriptor for shape pattern recognition. Aust J Basic Appl Sci 12(7)
7. Singh V, Kumar S, Dimri PK (2019) Comparative study of XPM-induced crosstalk due to 3OD parameter in SCM-WDM transmission system. Optik 186:177–181
8. Federici JF (1991) Review of four-wave mixing and phase conjugation in plasmas. IEEE Trans Plasma Sci 19(4):549–564
9. Hamdi MM, Mohammed MJ (2018) BER Vs Eb/N0 BPSK modulation over different types of channel. Aust J Basic Appl Sci 12(5)
10. Inoue K (1992) Four-wave mixing in an optical fiber in the zero-dispersion wavelength region. J Lightwave Technol 10(11):1553–1561
11. Watanabe S et al (2019) Wavelength conversion using fiber cross-phase modulation driven by two pump waves. Opt Express 27(12):16767–16780
12. Abood MS et al (2020) Fuzzy Logic Controller to control the position of a mobile robot that follows a track on the floor. In: 2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT). IEEE
13. Hamdi MM et al (2020) Coarse WDM in metropolitan networks: challenges, standards, applications, and future role. J Phys Conf Ser (IOP Publishing)

14. Liu Z-Y et al (2016) Large cross-phase modulations at the few-photon level. Phys Rev Lett 117(20):203601
15. Zheng Q et al (2019) XPM mitigation in WDM systems using split nonlinearity compensation. In: Asia communications and photonics conference. Optical Society of America
16. Nazemosadat E, Pourbeyram H, Mafi A (2016) Phase matching for spontaneous frequency conversion via four-wave mixing in graded-index multimode optical fibers. JOSA B 33 (2):144–150
17. Abood MS, Ismail M, Nordin R (2016) A quiz management system based on P2P near-field communication on Android platform for smart class environments. In: 2016 international conference on advances in electrical, electronic and systems engineering (ICAEES). IEEE
18. Huang T et al (2015) All-fiber multiwavelength thulium-doped laser assisted by four-wave mixing in highly germania-doped fiber. Opt Express 23(1):340–348
19. Raza N, Zubair A (2019) Dipole and combo optical solitons in birefringent fibers in the presence of four-wave mixing. Commun Theor Phys 71(6):723
20. Marcuse D, Chraplyvy AR, Tkach R (1991) Effect of fiber nonlinearity on long-distance transmission. J Lightwave Technol 9(1):121–128
21. Yang X, Hechao Y (2010) The application of optiSystem in optical fiber communication experiments. In: Proceedings of the third international symposium on computer science and computational technology (ISCSCT'10). Citeseer

Mobile 5G: Towards Technical View, Challenges and Issues

Ahmed Shamil Mustafa, Mustafa Maad Hamdi, and Mohammed Salah Abood

Abstract The need for networking infrastructure has greatly increased in the growing number of mobile devices. Higher data rates for new applications have pushed the existing technologies further forward. There were greater capacities and more effective deployments from 1G networks to 4G networks. However, the new 4G systems are now drained. With its latest technologies, 5G, the next version of networks, will overcome 4G problems, enabling mobile devices to be built again. We provide a theoretical perspective on technologies planned to be used in 5G systems in this article, including the addition of a brief review of previous mobile generations (1G to 4G) to illustrate 5G advantages. This paper also emphasizes the value of URLLC in line with technology and market requirements in the near period. The fields of study open to further analysis in the application of URLLC are outlined, and URLLC is successfully applied.

Keywords 5G · New mobile generation · High-speed networks · LTE · URLLC

1 Introduction

Wireless communication technology is one of the most important wireless access systems for core networks [1]. Mobile phones were only used to express voice in the early 80 s. With rapid technological growth, mobiles in the first and third generations have been revamped, and in 2009 a fourth version was finally launched. A growing generation has been introduced to further software and chat technologies, beginning with Internet surfing and not stopping with 3D video communication [2–4]. But the available resources hit their peak with the growing number of

A. S. Mustafa (✉) · M. M. Hamdi
Al-Maarif University College, Ramadi, Iraq
e-mail: ahmedshamil90@uoa.edu.iq

M. S. Abood
School of Information and Communication Engineering, Beijing Institute of Technology, Haidian District, Beijing 100081, China

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_30

technology users. One of the most important obstacles for wireless communication is the physical shortage of the Radio Frequency (RF) range [5]. Wireless frequencies normally between several hundred Megahertz and several Gigahertz have been used heavily, which is a big challenge to increase the number of system users. Another drawback that affects new wireless technology is high energy consumption. The increase in energy use in wireless communications networks implicitly results in an increase in CO_2 pollution, already considered a major environmental hazard that contributes a major contribution to electrical bills, thereby reducing its cost-efficiency. It is also a big challenge to provide communication with high data rates. 4G data networks and their current technologies have almost exceeded the potential data rates, and new technologies are needed to combat the previous problems [6–8]. 5G can provide a wider range of customers with high-speed data communications, power-efficient and high mobility base stations with its new technologies. In its network architecture, networking approaches and some services, 4G will be completely updated. According to the study, indoor and outdoor network access conditions are isolated and large MIMO networks are therefore updated to 5G within the 4G program. In this article, we present the technological viewpoint of the next generation of wireless communications (5G), summarizing and contrasting the key strategies implemented. Section 2, summarizing all four previous generations with their technical specifications, Sect. 3 provides a detailed technical perspective on 5G and its technologies. Eventually, Sect. 4 provides a challenges and issues. We finish this review paper with Sect. 5 which is include a technical comparison.

2 From 1G Towards 4G

2.1 First-Generation (1G)

The first mobile network generation came into being in early 1980. At the time, the network was called a cellular system and mobile phones were used to contact people. In the first generation networks, Advanced Mobile Phone Network (AMPS), Total access communications systems (TACS) and Nordic Cell Telephone (NMT) were among the mainstream [9]. As the number of cellular users grew, in 1990 GSM networks extended 1G to 2G.

2.2 Second-Generation (2G)

The second generation of 2G mobile technology was launched in 1990. Virtual modulation global communication network (GSM) for the enhancement of voice call efficiency has been introduced. In contrast with the first wave, transmission, and processing efficiency were improved. Certain systems contained text messages,

faxes, paging and voicemail [10]. GPRS was introduced to provide switch-package data to current GSM networks. In 2G, GSM technology, eight channels per network are served on the whole rate channel with a gross data ratio of 22.8 kbps [11]. Subsequently, GPRS increased the data rate to 64 kbps. Phone 2G services continue to be used extensively in many parts of the world.

2.3 Third Generation (3G)

In early 2000, the market for high-speed wireless connectivity exploded with the wide growth of smartphones and Internet applications. 3G networks of the third century have been designed to satisfy these requirements. UMTS and High-Speed Packet Service HSPA both provided high-speed networking facilities, cloud storage to multimedia software. 3G-based network speeds of 2 Mbps, with a 1.8–2.5 GHz frequency band, 3G-based access levels varying from satellite and rural outdoor 144 kbps, urban outdoor 384 kbps and indoor and outdoor low range [12, 13].

2.4 Fourth Generation (4G)

The next generation of wireless technology was launched in 2009 with a huge increase in mobile users. Since the performance levels of 2G and 3G are small, 4G was planned to provide data rates of up to 20 Mbps while offering improved quality and services. IP based packet switching was a good solution for connecting high-speed data. A system called IP-based networking has the capability to handle mobile and fixed networks at 100 Mbps and 1 Gbps data rates, respectively. 4G technology includes Long Term Growth, Enhanced LTE and Worldwide Microwave Communication Interoperability (Wi-MAX) [14]. Table 1 show a comparison of four groups of their characteristics and used techniques.

3 5G Technical View

Like all previous generations, 4G networks and new technologies will be launched from the fifth generation. New network architecture for 5G has been implemented. Indoor and outdoor conditions are separated to optimize efficiency and resource use. Indoor and outdoor, massive MIMO systems are designed to achieve greater efficiency as well as reduced power usage. More technologies can be applied to improve communications to short-range data particularly for indoor scenarios such as Visible Light Communications (VLC) and millimeter-wave technology. Nonetheless, other important elements such as scheduling strategies and network access standards will be omitted from 4G [15].

Table 1 Review of various mobile generations

Specification	1G	2G	3G	4G
Start	1980	1990	2000	2009
Peak bitrate	Not supported	64 kbps	2 Mbps	20 Mbps
Frequency band	800–900 MHz	850–1900 MHz	800 MHz–2.5 GH	800 MHz–3 GH
Multiplexing	FDMA	TDMA	CDMA	OFDMA
Standards	AMPS, TACS	GSM, EDGE, GPRS	UMTS, HSPA	LTE, WiMAX
Core network	PSTN	PSTN	Packet-switched	All IP networks
MIMO	No	No	Yes	Yes
IP support	No	Yes	Yes	Yes

3.1 5G Network Architecture

Since wireless operators remain indoors for about 80 percent of the time while remaining outdoors only around 20% [16], it is suggested to distinguish indoor and outdoor access networks, which decrease connectivity loss and thereby improve wireless transmission data speed, spectral efficiency, and energy efficiency. A distributed antenna network (DAS) that involves hundreds of antenna components and a large MIMO with broader antenna arrays [17] is used to separate interior and exterior scenarios.

For outdoor BSs, a large number of antenna arrays with a high number of antenna elements are placed around the cell in several locations, and these antennas are linked by optical fibers to the base stations. Through coordination between different user antennas and BS antennas, practically huge antenna arrays will be designed for the creation of vast virtual MIMO connections. With wireless connectivity points within the house, wide antenna arrays are transmitted through cables that definitely improve the average cellular system performance, spectral efficiency, energy efficiency, and data rate. In order to solve the problem of bandwidth depletion by non-traditional parts of the specter, various technologies can be implemented indoor for situations such as a femtocell, UWB, mm-wave (3–300 GHz) communications [10] and visible light communication (VLC) (400–490 THz) [18]. Such technologies will considerably increase the indoor scenario data rate to approximately 1 Gbps. 5G should be a heterogeneous network that provides service to all consumers across macrocells, microcells, picocells, and relays in order to achieve better performance. It is also necessary to accommodate high mobility users such as cars and train passengers. Such requirements are met by assigning a small cell to every mobile individual with a mobile femtocell concept [19]. The computational framework for mobile 5G is shown in Fig. 1. It is clear that both indoor and outdoor conditions are separated, also the femtocell concept is shown.

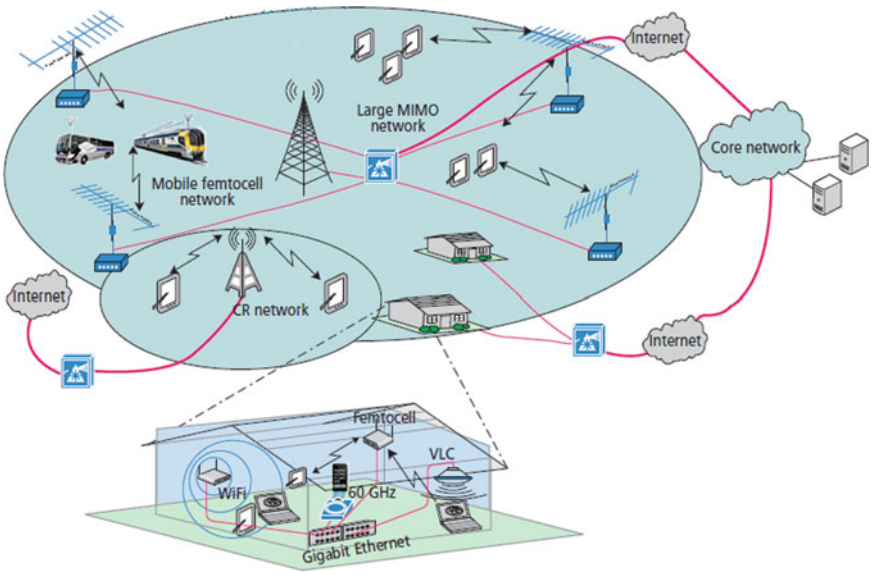


Fig. 1 5G Proposed model of network

3.2 4G Technologies Within 5G

Inside 5G networks, most 4G apps may continue to be used. The main radio technology will be the modulation of the OFDM and SOFDM frequency divisions using WiMAX and LTE requirements. Different links (SC-FDMA and MC-FDMA) will be introduced in conjunction with MIMO systems to provide wireless network access [20]. IPv6 packet-switched connections are used to serve additional applications and help all IPv4 users without Network Address Translation (NAT). The frequency range usable, 800 MHz–3 GHz used by 4G, is expanded to 8 GHz, offering wider bandwidth networks. The upstream and downstream divisions will be divided by the Duplexing Time and Frequency Divisions (TDD and FDD) in exactly the same way as in LTE and WiMAX. Figure 2 screens both OFDMA and SC-FDMA expected in 5G networks [16, 21].

3.3 5G New Technologies

In order to provide previously mentioned benefits, various new innovations will be introduced to 5G networks, some of them.

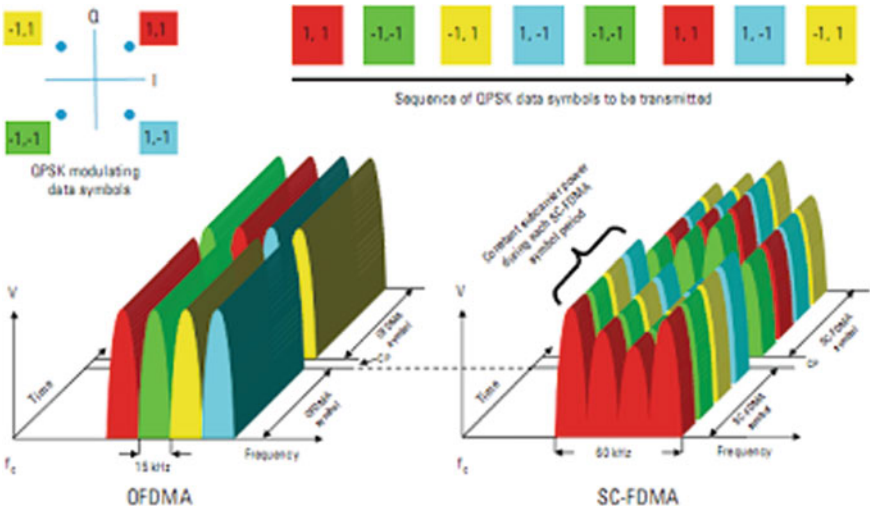


Fig. 2 OFDMA and SC-FDMA are used by 4G and are also planned for use in 5G

Massive MIMO Systems

The application of multiple antennas to both the transmitter and the receiver will achieve greater spectrum performance in a co-located or dispersed way (DAS distributed antenna systems). Even the antenna components can grow to tens or hundreds, adding more versatility and increasing the measurements of time and frequency. Each antenna is concurrently owned by many machines in this form of systems and therefore large MIMO systems deliver improved spectral efficiency and energy efficiency [22].

Femtocells

Mobile femtocells will be used in smart transportation systems of the next decade [19], and this design blends the mobile relay model with femtocell technology. The growing mobile individual is allocated to a femtocell, which can easily move and adjust its link to the different core networks of the operator. Such device is primarily used in public transport systems, such as taxis, trains and private cars. The use of femtocells could potentially improve the network's spectral efficiency. Femtocells will also increase the data rate by eliminating overhead signals required for the transmission as this task is carried out by femtocells for all its related applications, resulting in high usability. Eventually, the limited gap of interaction between femtocells will increase power efficiency, since each consumer only communicates with the femtocell base station [23].

Visible Light Communications (VLC)

In recent years, visible light encounters have been proposed. VLC can provide bandwidth and fiber optics-like data rates, making it a good replacement for RF communication. The signal transmitters can be used as white light-emitting diodes

(LEDs) used for illumination purposes [24]. VLC permits devices that activate and concurrently provide wireless broadband data connectivity. Infrasound (IR) LEDs or RF technologies such as Wi-Fi and Bluetooth will provide the uplink. Various modulation methods must be changed in compliance with the requirements of an optical signal, although many modulation schemes such as OFDM with discrete multitone modulation (DMT) have been written. Data speeds of up to 3.5 Gb/s were registered for one RGB LED [25].

4 Challenges and Issues

The biggest obstacle of 5G is to be implemented in a radically different IoT model attributable to a very large range of integrated networks, including essential infrastructure, with asymmetric resource measurements. To define preferred capabilities, the problem is radio security, computer security, network security, Denial of Service (DoS) attack protection; distributed control systems that include planning to avoid signaling storms; DoS attacks on end-user machines (operating systems, software and user data configuration) and the so on. In addition, because of the difficulties in restricting a successful attack, the attack will be more widely observed because of the cross- and intra-network borders stemming from the functional distinction between the underlying networks and processing facilities (NFVs and SDNs) and not from the physical separation [26–28].

The ITU categorizes the fifth (5G) generation of ultra-reliable low-latency communications (URLLC), mobile enhanced broadband and large-scale roll-calling (MMTC) communications in a manner that responds to numerous industrial and consumer demands. Studies have been carried out in order to ensure effective application within the usable scope of the aforementioned distributions [29].

The implementation of the URLLC specifications is also a challenge for 5G networks which would entail major improvements to the existing telecommunications service system architecture. A crucial position in real-time and mission-critical networks implementations. Implementing URLLC as mentioned below has several issues [29].

4.1 Quality of Service (QoS)

An end-to-end (E2E) distribution of data with reliability, protection and low latency needs URLLC-focused applications. The 3rd Generation Partnership Project (3GPP) has implemented specifications such as an air latency of 1 metric ms and 99,999% device efficiency for URLLC to carry out optimal Quality of Service (QoS) criteria [30].

4.2 URLLC Packet Design and URLLC Scheduling

One of the biggest challenges in URLLC is packet architecture. In terms of packet processing time and packet transmission time the delay can be reduced by an efficient packet configuration [31]. Packet processing requires time to buy a packet, chanson entry, scheduling (control) knowledge retrieval, packet encoding and correcting errors. One of the most critical problems is to program an unpredictable URLLC packet generation. User data is stored in a device-specific transfer buffer as user data arrives. No less than one TTI is required to transmit each packet. However, conditions on the radio link, payload size and resource availability can involve scheduling to boost the package TTI [32].

4.3 Energy Efficiency Concern for End-User Device

Many wireless systems use an energy-saving sleep mode. The devices must respond to any delay immediately after a packet is received from a network. Likewise, applications constantly track packets waiting for latency on the network. For URLLC-based operation, the existing energy saving states for User Devices (UE) are not suitable. However, due to the high level of data processing across the network, the EU will lose battery dramatically [29].

4.4 Handover Issues for URLLC and Error Handling

Handover is one of the most detailed elements of all telecommunication networks. Mobility criteria must be met by NR for 5G. The method of the simple transfer is very close to the method of the LTE transfer. In order to ensure smooth implementation, NR facilitates handover at two separate stages. Cell level mobility is controlled in the same way as with the LTE handover through a radio resource control (RRC) layer. The mobility of the beam is often treated with physical and medium-access (MAC) layers without the use of low latency RRC [29]. Because of the data channel's flawed nature, managing packet errors is yet another matter to be tackled along with URLLC latency, e.g. the 1 ms latency date. At the price of higher latency, this actually LTE gives a very low error rate, which does not suit URLLC [33].

Table 2 4G anf 5G networks A comparison

Specification	4G	5G
Launch date	2009	2015
Frequency band	800 MHz–3 GHz	800 MHz–8 GHz
IP technology	IPv6 + IPv4	IPv6 + IPv4
Core network	All IP network	Internet
MIMO implementation	4 × 4 MIMO for downlink and 2 × 2 MIMO for uplink	>10 × 10 MIMO for both uplink and downlink
Expected peak data rate	1 Gbps for fixed stations and 100 Mbps for mobiles	10 Gbps for fixed stations and 1 Gbps for mobile
Users capacity/ smallest cell	2 high capacity users	20 high capacity users
Latency	1–1000 ms	<1 ms
Smallest cell size	Few tens m^2	10 m^2

5 5G Technical Comparison

The distinction involves both the common and technological aspects as indicated by the authors of the model (Table 2). Resume the basic parameters for these two structures. This segment provides a comparison of 5G and 4G.

In the above example, it is obvious that 5G delivers better cell-sized usage efficiency than ten times greater. In fact, higher data levels will be reached by adding large numbers of MIMO components. New applications are thus easily supported as systems for virtual reality and five sensory communications.

6 Conclusion

In this article, we have presented a detailed study of 5G mobile communications technology. With the constant growth in the number of mobile devices, upgrading to 5G is important. 5G meets the growing bandwidth requirements due to its new technologies such as mass MIMOs and transparent lighting communications. According to new network configurations, increased power output and cost efficiency are anticipated from 5G. New applications, which with current generations are still undoable, can eventually find their way to everyday life. This review also deals with a few main areas of URLLC communication problems and concerns. Based on the problems discussed in the document, readers are also presented with a range of areas already available to further research in URLLC improvements.

References

1. Taleb T, Ksentini A, Kobbane A (2014) Lightweight mobile core networks for machine type communications. IEEE Access 2:1128–1137
2. Rowles D (2017) Mobile marketing: how mobile technology is revolutionizing marketing, communications and advertising. Kogan Page Publishers
3. Jasim NA, Abbood S, Salah M (2018) Speech synthesis using neural network. Open Sci J 3
4. Talab MA, Awang S, Ansari MD (2020) A novel statistical feature analysis-based global and local method for face recognition. Int J Opt
5. Mai TV, Molnar JA, Rudd K (2011) Security vulnerabilities in physical layer of cognitive radio. In: 2011 IEEE 54th international midwest symposium on circuits and systems (MWSCAS), pp 1–4. IEEE
6. Chen S, Zhao J (2014) The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication. IEEE Commun Mag 52:36–43
7. Abood MS, Thajeel IK, Alsaedi EM, Hamdi MM, Mustafa AS, Rashid SA (2020) Fuzzy logic controller to control the position of a mobile robot that follows a track on the floor. In: 2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT), pp 1–7. IEEE
8. Mohammed JM, Hamdi MM (2018) Hybrid tow feature extraction descriptor for shape pattern recognition. Aust J Basic Appl Sci
9. Frattasi S, Fathi H, Fitzek FH, Prasad R, Katz MD (2006) Defining 4G technology from the users perspective. IEEE Network 20:35–41
10. Kumar V, Kumar V (2013) A new generation wireless mobile network-5G. Int J Comput Appl 70
11. Joshi NC (2016) Mobile communication-5G technology. Int J Res Humanit Soc Sci 4
12. Goyal P, Buttar AS (2015) A study on 5G evolution and revolution. Int J Comput Netw Appl (IJCNA) 2
13. Abood MS, Ismail M, Nordin R (2016) A quiz management system based on P2P near-field communication on Android platform for smart class environments. In: 2016 international conference on advances in electrical, electronic and systems engineering (ICAEES), pp 83–88. IEEE
14. Khan J, Abbas A, Khan K (2011) Cellular handover approaches in 2.5 G to 5G technology. Int J Comput Appl 21:0975–8887
15. Mumtaz S, Huq KMS, Rodriguez J (2014) Direct mobile-to-mobile communication: paradigm for 5G. IEEE Wirel Commun 21:14–23
16. Wang C-X, Haider F, Gao X, You X-H, Yang Y, Yuan D, Aggoune HM, Haas H, Fletcher S, Hepsaydir E (2014) Cellular architecture and key technologies for 5G wireless communication networks. IEEE Commun Mag 52:122–130
17. Tang W, Chen MZ, Dai JY, Zeng Y, Zhao X, Jin S, Cheng Q, Cui TJ (2020) Wireless communications with programmable metasurface: new paradigms, opportunities, and challenges on transceiver design. IEEE Wirel Commun
18. Haas H (2011) Wireless data from every light bulb. TED website 49
19. Haider F, Wang C-X, Haas H, Yuan D, Wang H, Gao X, You X-H, Hepsaydir E (2011) Spectral efficiency analysis of mobile femtocell based cellular systems. In: 2011 IEEE 13th international conference on communication technology, pp 347–351. IEEE
20. Banchs A, Gutierrez-Estevez DM, Fuentes M, Boldi M, Provvedi S (2019) A 5G mobile network architecture to support vertical industries. IEEE Commun Mag 57:38–44
21. Hamdi MM, Mohammed JM (2018) BER Vs Eb/N0 BPSK modulation over different types of channel. Aust J Basic Appl Sci
22. Rusek F, Persson D, Lau BK, Larsson EG, Marzetta TL, Edfors O, Tufvesson F (2012) Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process Mag 30:40–60

23. Lu Y, Jia X, Chen Y (2020) Spectral efficiency analysis of millimeter-wave heterogeneous cellular networks based on PCP. J Phys Conf Ser 012120
24. Rehman SU, Ullah S, Chong PHJ, Yongchareon S, Komosny D (2019) Visible light communication: a system perspective—overview and challenges. Sensors 19:1153
25. Slaiman I, Boon T (2014) Optical wireless communications through visible light leds: an overview. Int J Eng Res Technol (IJERT) 3
26. Bhardwaj A (2020) 5G for military communications. Procedia Comput Sci 171:2665–2674
27. Maad Hamdi M, Audah L, Abduljabbar Rashid S, Al-Mashhadani MA (2020) Coarse WDM in metropolitan networks: challenges, standards, applications, and future role. J Phys Conf Ser 1530:012062
28. Mustafa AS, Hamdi MM, Mahdi HF, Abood MS (2020) VANET: towards security issues review. In: 2020 IEEE 5th international symposium on telecommunication technologies (ISTT), pp 151–156. IEEE
29. Siddiqi MA, Yu H, Joung J (2019) 5G ultra-reliable low-latency communication implementation challenges and operational issues with IoT devices. Electronics 8:981
30. Li C-P, Jiang J, Chen W, Ji T, Smee J (2017) 5G ultra-reliable and low-latency systems design. In: 2017 European conference on networks and communications (EuCNC), pp 1–5. IEEE
31. Ji H, Park S, Yeo J, Kim Y, Lee J, Shim B (2018) Ultra-reliable and low-latency communications in 5G downlink: physical layer aspects. IEEE Wirel Commun 25:124–130
32. Pocovi G, Pedersen KI, Mogensen P (2018) Joint link adaptation and scheduling for 5G ultra-reliable low-latency communications. IEEE Access 6:28912–28922
33. Shariatmadari H, Li Z, Iraji S, Uusitalo MA, Jäntti R (2017) Control channel enhancements for ultra-reliable low-latency communications. In: 2017 IEEE international conference on communications workshops (ICC Workshops), pp 504–509. IEEE

Federated Learning Approach for Privacy-Preserving on the D2D Communication in IoT

Tanweer Alam

Abstract Federated Learning (FedL) allows smart devices to use the collective predictive models efficiently and effectively when having most of the trained data on the device, essentially decoupling advanced analytics for the requirement to maintain information in the database. The Device-to-Device (D2D) messaging on the Internet of Things (IoT) is characterized as immediate interaction among smart devices without going through the access point or the sensor nodes. D2D approach is usually never visible to the base station. Throughout a conventional communication system, all transactions should be carried out throughout the base stations even if the correspondence participants are beyond the range of the networks. The IoT devices enable communication with each other and access throughout the IoT system. The integration of existing communications networks including IoT systems offers a wide range and advanced mobile networking services. The message scheduling algorithm operates on the distributor basis to take scheduling decisions on the communicated messages. Such a schedule enhances the performance of the D2D communication. Through the implementation of FedL, the author assumes a system of wireless connections shared a public faded communication link. Where every device has its training collection, and connectivity is usually accomplished through D2D networking. In this article, the author has proposed a D2D secure communication algorithm for IoT devices. This algorithm is tested using some experiments and it performs well.

Keywords Federated learning · Device-to-device · Wireless communication · Internet of things (IoT) · Smart devices

T. Alam (✉)
Faculty of Computer and Information Systems, Islamic University of Madinah, Madinah, Saudi Arabia
e-mail: tanweer03@iu.edu.sa

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_31

1 Introduction

FedL is an extremely novel approach to train that eliminates the use of central control information gathering through the model training process. Information is obtained through various channels (e.g., IoT devices) and processed inside a centralized point in a typical artificial intelligence process. When all the information is centralized, a unified machine learning model is trained using it. Such a method has been known as distributed training so the information should be transferred from the participants' device to a centralized database for the model to be designed and trained. FedL provides several significant strengths over centrally controlled learning algorithms:

1. Because the information is stored on the local machine that ensures privacy.
2. Since the reconfigured approach could be utilized to produce accurate predictions on the local device, there is reduced latency.
3. Considering the collective learning method, so that the smarter models are possible.
4. Since models are trained on a local device so that there is reduced energy usage required.

The IoT system represents a network of interconnected things which are integrated with sensor devices, programs, as well as other techniques for the way of transferring and sharing information between devices and services within the system. A huge number of smart devices, Radiofrequency identification, sensors, and actuators could be interconnected through the IoT access point [1]. A huge number of transactions produced by every object is typically low, it does not need to be consistent, however, the cumulative amount of data could be very huge. Latest predictions expect that a huge proportion of this kind of data would be moved into the wireless communication network, which would be one of the main reasons behind the upcoming fifth generation (5G) networks [2]. The strengths of D2D interactions could be divided into two main categories: (i) improving network capacity and (ii) allowing new location providers. The D2D technology improves system throughput through enhancing energy and/or bandwidth quality, reducing latency, avoid congestion, expanding wireless coverage and handling network activity, and so on [3–5] (Fig. 1).

Throughout fact, community security agencies and various private and public presence resources are made available via D2D interaction. This advent of 5G technology has contributed to a revision of D2D communications from a 5G technology perspective. Recent research developments in D2D interaction are emerging, in particular: vehicle-to-vehicle (V2V) interactions, V2V transmitting for the Web of Things and fitness trackers, infrared imaging technologies, and culture security systems [6–8]. The D2D transceiver shall be combined in multipath channels. Synchronous replication is given by the system in the situation of D2D convergence. However, alignment among D2D communication is not permitted because D2D pairs devices which have specific scheduling advanced changes (for

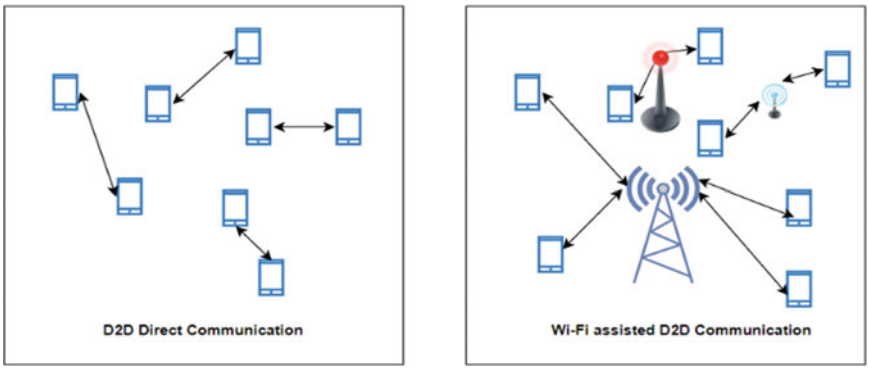


Fig. 1 D2D direct and network-assisted communication

example whenever devices are connected to two non-synchronous cells or even when devices are in the same cells however at varying distances [9–11] (Fig. 2).

The limitation would be enhanced by the use of D2D methodology, whereby D2D networks are measured at the stage of the D2D recipient and instead communicated to the intermediate nodes. Through this follows the concept of providing decentralized scheduling which enhances D2D consumers' awareness of the specific communication system. Its main goal of the proposed schedule system is to reduce energy usage and boost system efficiency. But we must point out that the proposed algorithms are relevant to every other success measure [12].

The rest of the paper is organized as follows. Section 2 represents the related work, Sect. 3 represents the methodology, Sect. 4 represents the results and discussion, Sect. 5 shows the conclusion.

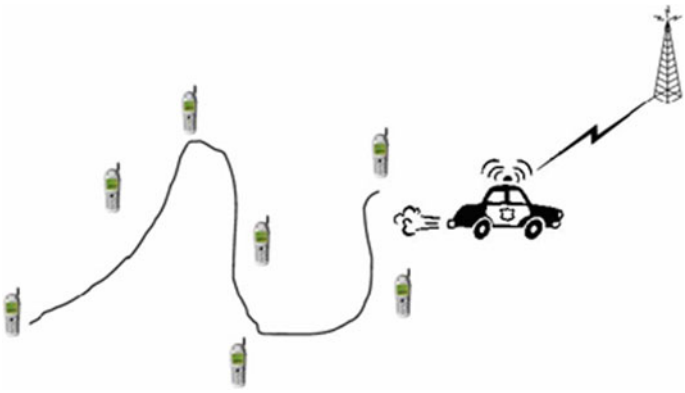


Fig. 2 IoT-based devices movements in the network range

2 Related Works

D2D messaging serves as a feasible model to enable the intercommunication of heterogeneous devices envisaged by IoT. The IoT framework could be a set of connected nodes that enable (through various computation and information systems, techniques, and methods) on a cooperative basis to improve quality. The number of smart objects is estimated to boost in the coming years, not only due to improvements in key features, and attributable to a decline in the cost of the gadgets and the increasing use of technology to improve the functionality of the devices at minimal expense. IoT is also recognized to be one of the primary technological solutions for the finally 5G wireless communication [13]. The D2D messaging can be used to enable fog computing endpoints applied in Long-Term Evolution (LTE) systems next to IoT gadgets. Primarily, multichannel area communications, which is commonly used by IoT apps to identify objects in nearness, could be introduced via D2D conversations. This helps the operator to conserve power and reuse energy. Throughout addition, researchers have suggested a transfer message enhancement that still optimizes D2D communications to improve LTE usability with a variety of device fog structures. The proposed development would include the investigation of several other main issues in the LTE architectural design proposed. Besides example, LTE enables Nodes to save more energy via irregular transmission. Researchers are looking at itself but how this strategy could be used to alleviate the issue of fuzzy clusters in IoT systems. Throughout addition, researchers are studying cross-layer efficiencies among network-and application-layer architectures [14] (Fig. 3).

In 2020, Yang Zhao et al., were published an article on the blockchain-oriented FedL approach for privacy-preserving. They were presented initially, the structured social networking FedL structure for building the classification model to assist household appliances producers in improving operational performance and

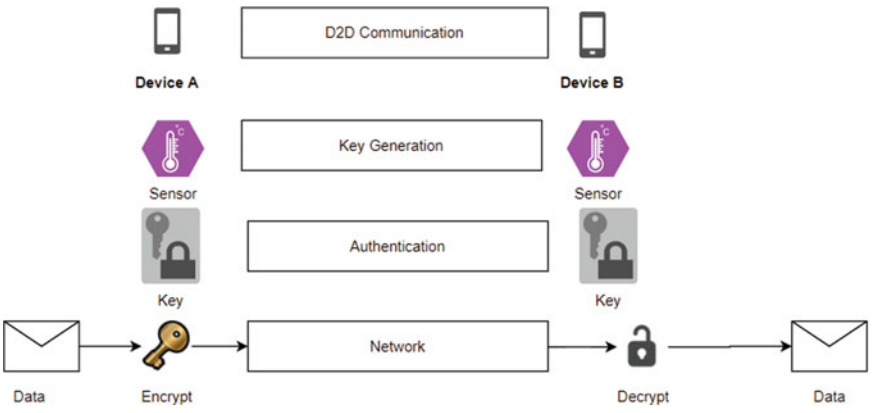


Fig. 3 D2D secure communication

optimizing household appliances functionality. Moreover, they suggest a novel simulation approach that is more accurate than batch classification although maintaining the confidentiality of each user's datasets. Furthermore, they restrict attackers from accessing the training dataset to predict confidential data about individuals by using evolutionary algorithms. Finally, their blockchain-based framework guards against unauthorized system modifications by requiring accountability for certain model upgrades [15].

Adrian Nilsson, et al., were published an article on the performance evaluation of FedL algorithms. In this paper, they have presented the measurement of the efficiency of several FedL techniques using a hierarchical procedure in which information is stored on the cloud. Also, on neural networks, the optimization techniques such as Federated Aggregation, Federated Probabilistic Variability Minimized Gradation, and other algorithms are evaluated [16] (Fig. 4).

Throughout the study, researchers suggested a two-tier 5G communication infrastructure with so-called microcells and third-party applications. Mostly in the implementation of design and simulation, the user may have various measures of access based on the economic approach provided. They also presented a classification of D2D communication, considering the level of control of the mobile controller. They also identified several of the significant technical issues that need to be handled for each form, like protection, interruption performance, and resource allocation matters.

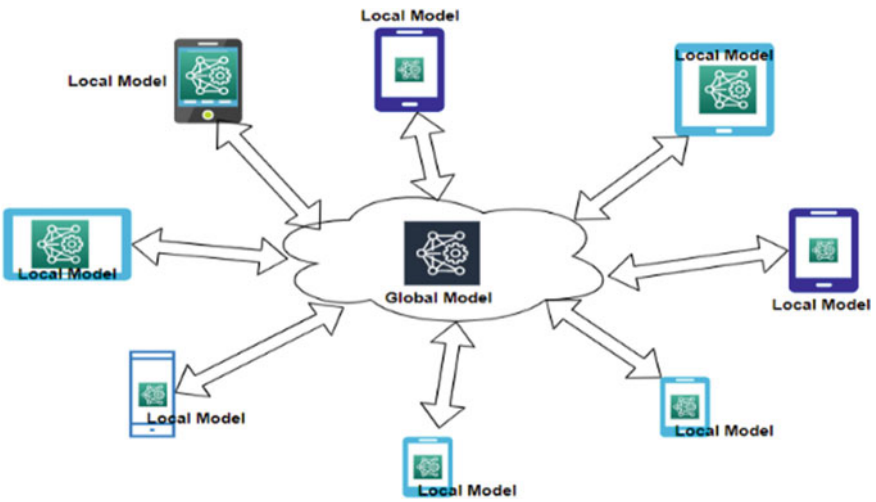


Fig. 4 Local and global models used in federated learning

3 Methodology

FedL for smart IoT technologies usually involves a centralized integration service (e.g., cloud server and wireless connection) and a set of IoT gadgets. Where every IoT device train the model on the device, although the integration service manages and improves the standard model stored on the cloud server. The standard and local models are improved recursively across a series of contact stages until the training accuracy achieves convergence [17] (Fig. 5).

Throughout each training phase, the three main operations are repeated. (1) The accumulation service just identifies a set of IoT gadgets for the localized training phase and upgrading to minimize the traffic load. (2) Local models attach: every IoT device gets updated the local model using optimization techniques focused on one of the latest frequently updated global models, then transmits the latest updates to the allocation service via mobile communication links for standard model development. (3) Standard model access: the cloud server integrates the local models from the chosen IoT gadgets to access the standard model, which is being transmitted through transmitting power exchanges to the designated IoT devices for the next phase of training.

The following approaches are used in D2D communication to recognize and discover IoT devices.

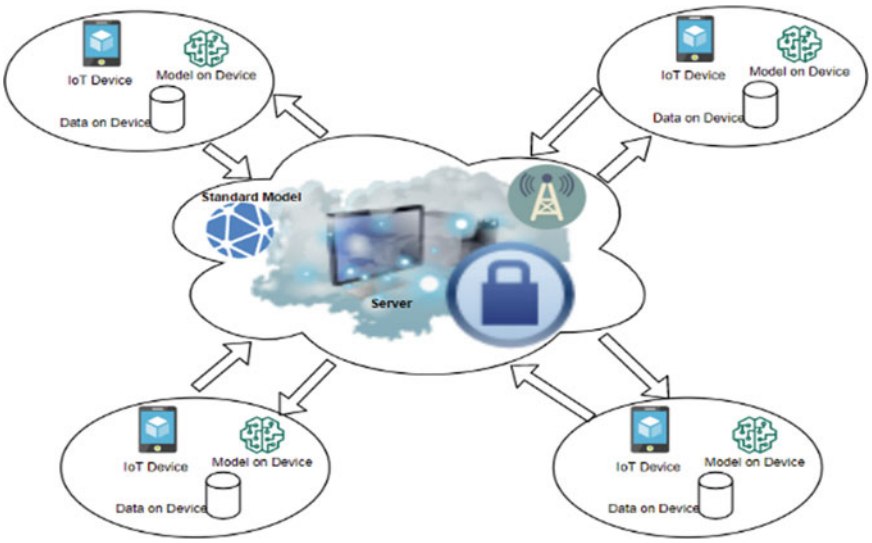


Fig. 5 Federated learning approach in D2D

3.1 Identify IoT Devices Utilizing the Probability-Based Approach

Suppose there is a D2D structure in the probability-based model. Network devices become usable across the x-axis, y-axis, and z-axis orientations in a 3-dimensional setting. Across the communication system, the entire region is separated through clusters. The area among all cells is configured in such a way that connected devices could travel inside the area of the cell. The IoT devices can find neighborhood devices in binary numbers inside the same cell region. An allowed IoT device has restricted the information which the other device discovers. The expectation of the practical exposure which goal movements could only happen among surrounding areas. Throughout fact, information is distributed using a weighted average attitude and the likelihood of motion. That attitude results from the versatility of an IoT system that monitors the direction saves the target area and establishes the motion.

3.2 The Discovery of Smart Devices Using the Hidden Markov Chain

The hidden Markov method is being used for the discovery of IoT devices in the planar region. This method is related to the workplace environment and the devices were traveling inside the zone and even this method can be found inside the context of nearby IoT nodes. Across the wireless communication area, we are forming a linear model, finding well all IoT devices and placing them in a convergence chain.

3.3 The Discovery of IoT Devices Using a Gradient Approach

The discovery of IoT devices using a gradient approach is designed to classify gadgets and exchange data to generate and submit additional data for the discovery of IoT nodes. It is not because of communication among Mobile connected devices, or simply because of the simplicity of the IoT node inherent in the D2D, such pattern remains. A gradient value picks 1 would also detect IoT devices in the area in which the IoT network is formed at such a stage where an IoT device identifies the destination. The gradient information is the place that has been based on moving through any distance.

IoT devices can be expressed by six tuples {I, S, T, L, Sp, O}.

Here I is the unique identification of IoT devices, S is the status of IoT node, T is the type of IoT nodes, L is the location of IoT node, Sp is the specification and O is the application objects.

All IoT devices have a unique identification number that is generated by the IoT network after entering the device into the network ranges. Suppose Device1, Device2 and Device3 are three IoT devices, connected in the network with the shared hash code H, the encryption key for Device1 is Device1E and the Decryption key for Device1 is Device1D. Assume that Device1 is sending information to DEVICE2 and DEVICE3. It can be represented as follows.

(1) Device1 → Device2: A, H (Device1E).

Device1 → Device3: A, H (Device1E).

DEVICE1 encrypts the DEVICE1E key along with hash code H and sends this information to DEVICE2 and DEVICE3.

(2) DEVICE2 and DEVICE3 decrypts H(DEVICE1E) and extracts DEVICE1E. DEVICE2 and DEVICE3 will generate 'ACK', encrypts it with H using the DEVICE1E key, and sent it to DEVICE1.

DEVICE2 → DEVICE1: H (DEVICE1E (ACK)).

DEVICE3 → DEVICE1: H (DEVICE1E (ACK)).

DEVICE1 received the information authenticates by DEVICE2 and DEVICE3, since DEVICE2 and DEVICE3 can access encryption key DEVICE1E from data sent by DEVICE1 to DEVICE2, DEVICE3. However, DEVICE2 and DEVICE3 should know the hash code H.

(3) DEVICE1 decrypts data sent by DEVICE2 and DEVICE3 and sends the ACK. DEVICE1also creates the challenge (ch)DEVICE1 and session SDEVICE1, then encrypts it using an ACK and sends it to DEVICE2, DEVICE3.

DEVICE1 → DEVICE2: ACK ((ch)DEVICE1, SDEVICE1).

DEVICE1 → DEVICE3: ACK ((ch)DEVICE1, SDEVICE1).

DEVICE1 sends ACK ((ch)DEVICE1, SDEVICE1) to DEVICE2 and DEVICE3. DEVICE2 and DEVICE3 can access ACK with hash code H.

(4) DEVICE2 decrypts received acknowledgments, access (ch)DEVICE1 and SDEVICE1. DEVICE2 generates hash function h((ch)DEVICE1) and ((ch) DEVICE2, SDEVICE2). So, the complete hash function (h ((ch)DEVICE1), (ch) DEVICE2, SDEVICE2) send to DEVICE1, encrypted by ACK acknowledgment.

DEVICE2 → DEVICE1: ACK (h ((ch)DEVICE1), (ch)DEVICE2, SDEVICE2).

DEVICE3 decrypts received data, access (ch)DEVICE1 and SDEVICE1. DEVICE3 creates hash function h ((ch)DEVICE1) and ((ch)DEVICE3, SDEVICE3). So, the complete hash function (h ((ch)DEVICE1), (ch)DEVICE3, SDEVICE3) send to DEVICE1, encrypted by ACK acknowledgment.

DEVICE3 → DEVICE1: ACK (h ((ch)DEVICE1), (ch)DEVICE3, SDEVICE3).

(5) DEVICE1 decrypts this information, extracts ((ch)DEVICE2, SDEVICE2) and ((ch)DEVICE3, SDEVICE3). DEVICE1 processes the hash function send by DEVICE2 and encrypts ACK acknowledgment. It sends this information to DEVICE2. DEVICE1 also processes the hash function send by DEVICE3 and encrypts the ACK acknowledgment. DEVICE1 send it to DEVICE3.

DEVICE1 → DEVICE2: ACK ((ch)DEVICE2).

DEVICE1 → DEVICE3: ACK ((ch)DEVICE3).

This shows the acknowledgment for DEVICE2 and DEVICE3 sent by DEVICE1.

Every IoT node DEVICE1, DEVICE2, and DEVICE3 knows the sessions SDEVICE1, SDEVICE2, and SDEVICE3. These devices can access the key (SDEVICE1, SDEVICE2, SDEVICE3) for sessions generated by DEVICE1. Now, they can start communication among DEVICE1, DEVICE2, and DEVICE3.

In the experiment, three devices are considered to evaluate the above procedure, However, in the future, we can enhance it for several IoT nodes. After generating the session, the IoT nodes can communicate directly. IoT nodes DEVICE1, DEVICE2, and DEVICE3 have used the challenges (ch)DEVICE1, (ch)DEVICE2, and (ch)DEVICE3 for confirming the secret hash code H generated by the blockchain. Al the end of the above procedure the IoT nodes will generate the final session key S = (SA, SB, SC) and distribute it to all connected nodes in the network.

An IoT node is enabled to collect data on the network. IoT is becoming an innovation of significant impact throughout many market sectors. Most IoT systems are designed to have worldwide coverage through thousands of easy and often embedded systems. Also, the limited capacities of many IoT devices, and the existing communications systems focused on centralized frameworks, are posing major issues in the IoT environment. Computational innovations are becoming increasingly important for facilitating smart IoT solutions including self-driving vehicles, unmanned aerial vehicles, and medical care, through extracting knowledge from information. The information acquired or obtained by a broad range of devices must be delivered to a centrally managed cloud storage, that can train a machine learning algorithm, through traditional fog artificial intelligence. Attempting to move Artificial intelligence from cloud storage to edge devices has received significant scientific and corporate initiatives due to the growing computing capacity of IoT devices and the growing consideration about data protection and privacy.

4 Results and Discussion

D2D communication is one of the key techniques offered to boost 5G efficiency. The D2D is a straightforward relationship among the two devices with no control center being involved this is suggested to increase performance, reliability, energy consumption, and bandwidth availability in IoT. Such an article describes an extensive optimization of the introduction of D2D connectivity into the wireless connection, concentrating on possible benefits, numerous recent experiments, categories, and implementations for D2D innovation. IoT devices are located inside the scope of D2D communication that embraces Wi-Fi access and connectivity. Every IoT process is designed to get a fixed Wi-Fi area and a revised messaging selection. The objective is to meet those criteria in terms of network coverage or conversation. As regards the organization of connected devices represented in an

analytical environment, the challenge is to find if the surface is occupied by k within the accurate region controlled by Wi-Fi in every situation when k is the success rate given. The D2D functionality framework is an optimized solution processing and D2D technology framework. The usability of D2D also depends on the versatility of devices and development platforms, including space and efficiency. Service providers establish digital communication capabilities, storage areas, and application interfaces that promote reliability, productivity, and integration. Connected devices from D2D can connect in the cloud convergence environment, although at most each IoT device should be connected to mobile or Wi-Fi networks. Each IoT device must be independently enrolled throughout the server.

The authors introduce the FedL-based mechanism for maintaining privacy for IoT devices throughout this article. The framework is built using a variety of cutting-edge technology, along with IoT devices, cloud services, and FedL. Furthermore, to preserve the security of the IoT device's data, the framework implements differential privacy. To strengthen the reliability of the FedL framework, authors develop a novel technique that performs admirably batch normalization when the privacy of functionality is preserved. Figure 6 shows that the training time taken on IoT devices versus cloud servers in seconds. Figure 7 represents the accuracy-test results through different experiments.

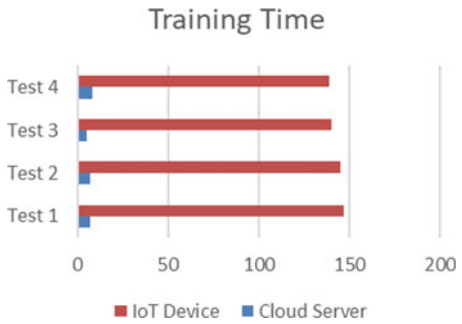


Fig. 6 Training time in seconds on the IoT device versus cloud server

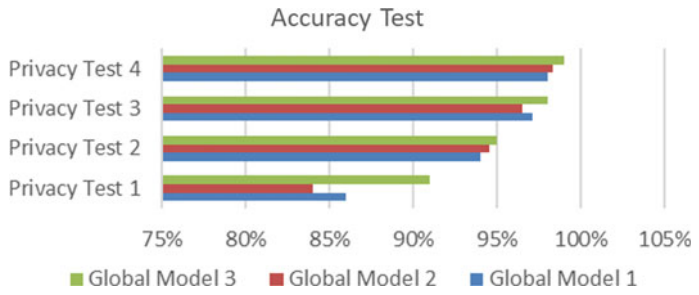


Fig. 7 Evaluate the accuracy using four tests

Automated computation, connectivity, and information sharing between devices have an important role to play in shaping the digital age. Evolving D2D networking does have a clear potential to promote peer-to-peer messaging. The D2D communication network is also considered to be an integral part of the IoT system. The massive rise in the network performance requirement is leading to major aspects of the design of the IoT framework. The proposed study shows the benefits and issues of smart D2D communication which will be established to achieve the IoT specifications. Several sectors and infrastructure providers have also shown considerable interest in the concept of the D2D methodology in wireless communication.

5 Conclusion and Future Scopes

Throughout this article, the authors describe a federated learning method for maintaining privacy in D2D communication on the Internet of Things. They measure the moment consumed by an IoT device versus a cloud server, as well as the accuracy test, from both a descriptive and analytical perspective, and use a consistency measure. Under certain multi-hop communication sensor networks, the strategies presented in the literature are not successful and/or realistic, according to our observations. This solution allows functioning without the critical monitoring of centrally controlled command, which increases the coverage of wireless connections and makes them highly efficient and productive when transferring data. Interconnecting devices provide different settings to produce IoT systems. The proposed D2D algorithm works well to schedule the messages in real-time. This scheduling technique boosts the performance of IoT networks. In the future, this approach can be implemented in IoT networking.

References

1. Alam T, Aljohani M (2015) An approach to secure communication in mobile ad-hoc networks of android devices. In: 2015 international conference on intelligent informatics and biomedical sciences (ICIIBMS), pp 371–375. IEEE
2. Alam T, Benaida M (2018) CICS: Cloud–internet communication security framework for the internet of smart devices. Int J Interact Mobile Technol (iJIM) 12(6):74–84
3. Shen X (2015) Device-to-device communication in 5G cellular networks. IEEE Network 29(2):2–3
4. Hisham M, Elmogy A, Sarhan A, Sallam A (2020) Energy efficient scheduling in local area networks. Wireless Netw 26(1):685–698
5. Liu T, Lui JCS, Ma X, Jiang H (2018) Enabling relay-assisted D2D communication for cellular networks: algorithm and protocols. IEEE Internet Things J 5(4):3136–3150
6. Shah ST, Hasan SF, Seet B-C, Chong PHJ, Chung MY (2018) Device-to-device communications: a contemporary survey. Wireless Pers Commun 98(1):1247–1284

7. Niu Y, Liu Y, Li Y, Chen X, Zhong Z, Han Z (2017) Device-to-device communications enabled energy efficient multicast scheduling in mmWave small cells. IEEE Trans Commun 66(3):1093–1109
8. Choi J, Park J (2020) SecureLinQ: joint precoding and scheduling for secure device-to-device networks. IEEE Wireless Commun Lett
9. Adnan MH, Ahmad Zukarnain Z (2020) Device-to-device communication in 5G environment: issues, solutions, and challenges. Symmetry 12(11)
10. Alam T, Benaida M (2018) The role of cloud-MANET framework in the internet things (IoT). Int J Online Eng 14(12)
11. Abdullah S, Asghar MN, Ashraf M, Abbas N (2020) An energy-efficient message scheduling algorithm with joint routing mechanism at network layer in internet of things environment. Wireless Pers Commun 1–15
12. Priya MD, Suganya T, Malar ACJ, Dhivyaprabha E, Prasad PK, Vishnu Vardhan LR (2020) An efficient scheduling algorithm for sensor-based IoT networks. In: Inventive communication and computational technologies. Springer, Singapore, pp 1323–1331
13. Suraci C, Pizzi S, Iera A, Molinaro A, Araniti G (2019) delivering multicast content through secure D2D communications in the internet of things. In: International conference on wired/Wireless internet communication. Springer, Cham, pp 182–193
14. Virdis A, Vallati C, Nardini G, Tanganelli G, Stea G, Mingozzi E (2018) D2D communications for large-scale fog platforms: enabling direct M2M interactions. IEEE Veh Technol Mag 13(2):24–33
15. Yang Q et al (2019) Federated machine learning: concept and applications. ACM Trans Intel Syst Technol (TIST) 10(2):1–19
16. Nilsson A, Smith S, Ulm G, Gustavsson E, Jirstrand M (2018) A performance evaluation of federated learning algorithms. In: Proceedings of the second workshop on distributed infrastructures for deep learning, pp 1–8
17. Xing H, Simeone O, Bi S (2021) Federated learning over wireless device-to-device networks: algorithms and convergence analysis. arXiv preprint: arXiv:2101.12704

A Performance Evaluation of Voice Over IP Protocols (SIP and H.323) in Wireless Network

Mustafa Sabah Noori, Ratna Kalos Zakiah Sahbudin, Mohammed Salah Abood, and Mustafa Maad Hamdi

Abstract Many types of Voice over Internet Protocol (VoIP) are used to provide audio or multimedia sessions over Internet Protocol (IP) networks, multimedia communications became very easy by using applications that provide Internet Protocol VoIP like Skype, Viber, WhatsApp, etc. And network simulation has frequently been used to determine the performance feasibility of any network protocols and the model reliability before practical implementation. Several kinds of studies were carried out on voice efficiency over different types of networks environment, among these experiments and researches, finding a better signaling protocol for communication by evaluating and analyzing the performance of the protocols, such as the most commonly used Session Initiation Protocol (SIP) and (H.323) signaling protocols. This paper provides a comparative study of the performance of the two aforementioned protocols that are most widely used over the IEEE 802.11a wireless network by using the QualNet simulator. The study was performed by analyzing the data packets for both initiator and receiver for both protocols, to evaluate the performance based on five metrics: Session total Bytes, Establishment time, Sent, total Bytes received of Real-time transport protocol (RTP), Average End to End Delay and Overall Throughput. The outcomes demonstrated SIP protocol ascendingly rise and superior over H.323 protocol as a permanently started from the Zero-point around 5.4 (Mbps) of the time until the end of the experiment close to 6.2 with a 0.05-s delay while the rival was always below these values of throughput. Hence its superior provides preferability in multimedia sessions. studying and evaluate the performance of two IPs with a high capability

M. S. Noori (✉) · R. K. Z. Sahbudin
Department of Computer and Communication Systems Engineering, Faculty of Engineering, Universiti Putra Malaysia, Selangor, Malaysia
e-mail: ratna@upm.edu.my

M. S. Abood
School of Information and Electronics Engineering, Beijing Institute of Technology, Beijing, China

M. M. Hamdi
Department of Computer Engineering Techniques, Al-Maarif University College, Al-Anbar, Iraq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_32

simulator QualNet, SIP outright outperforms H.323 provide preferability recommendations in voice calls.

Keywords QualNet · Simulation · SIP · H.323 · VoIP · Wireless network · Signaling protocols

1 Introduction

Millions of customers communicate every day across the internet and the exponential growth and success of various media communications lead to a produce vast amount of data like audio, video, text, etc., and for the last decade, many communications based on VoIP protocols have been widely used for multimedia large data communications [1]. VoIP services are growing increasingly due to their versatile and straightforward implementation and the low cost of traditional public telephone networks [2]. VoIP is a standard instant message chat technology and commonly used in social networking, has the potential to deliver instant message services with less capital, and is exceptionally scalable [3]. This fact underlines the significance of voice-over IP signaling protocols. For many years, accordingly, two protocols have competed in this field, namely H.323 and SIP [4]. They came from various organizations. H.323 is the product of the ITU Telecommunication Standardization Sector (ITU-T), which has been the most relevant agency in telecommunications and telephony for a long time. In contrast, SIP comes from the Internet Engineering Task Force (IETF) this organization oversees the growth of the Internet [5]. We will provide a study and evaluation of these two IPs in terms of properties and performance and communications behavior even in Quality of Service (QoS) [6], and give details on why one was more effective in this mission, the evaluation characteristics are functions of network components, applied applications in the protocol architecture, and supported applications for networking in addition to peer-to-peer telephony [7].

However, the definition of the user agent back-to-back is significant in the creation of complex SIP network applications [4]. Back-to-back users include two user agents, who are defined somewhat as the most complicated form of a SIP proxy, one uses the incoming session, and the other user agent sets another session. Technology at the framework level monitors two sessions signaling and transmits signals from one to the other (after making necessary modifications to messages). Enables specific programs to be introduced, which need some session control and media stream. In a collection of Request for Comments RFC documents [5]. Protocol specification of the SIP provides. Key RFC Paper defines only the portion and features of the protocol needed for first-party call power. Separate records include case notifications mechanisms, third-party call management elements, the presence, instant message services elements, meeting features, and various other elements of the protocol. Any of them are going to be listed here: The main documents, the generic event of notification machine, notification of presence,

Notice of Registration, Presence Information Data Format (PIDF), model for the development of presence and instant messaging applications and Call Processing Language (CPL), message request types.

On the other side, our second scope protocol is H.323 protocol is ITU Telecommunications (ITU-T) standard recommendations are that specified for any packet network to provide audio-visual communication sessions. The H.323 protocol describes multiple connectivity components such as Gateways, Gatekeepers, Multipoint Control Units, and Internet-based and Boundary Elements that provide aggregation and review of multimedia connectivity capabilities. H.323 has seven steps of a call which will facilitate the exchanging of messages or supplementary resources between components. The following are six essential phases of the H.323 functionality: setup call, communication initialization, and exchange capability, visual audio communication establishment, call stability with RTP, media service exchange on call, and termination the call. Although the protocol H.323 has more architectural complexity, it offers continuity of control by schedule calling and has an efficient transmission that allows the system more robust and secure [8]. H.323 is a group of protocols coupled with each other to ensure three primary tasks: signaling, negotiation of codecs, and transport of VOIP data. the protocol guarantees control over the quality and usage of services constraints. Its protocol stack consists primarily of the first H.245 protocol that negotiates the opening and usage of channels and the configurations for VOIP contact. The second Q.931 protocol enables the establishment of signals and calls. Third, the protocol of enrollment, admission, status 'RAS' is the protocol used for the terminal to interact with the gatekeeper. Last Usage of the Real-Time Streaming Protocol (RTCP)/RTP protocol to route and manage sources of audio and video [9].

Besides the general specifications for H.323, there are additional specifications [10]. H.225 and H.24 are the two most important, but the list does not end here, since there are H.235 specifications (which specify H.323 security functions) as well as other specifications. Because the communication network simulators are limited abilities, using the QualNet network simulator is considered to be more effective in terms of network scalability and accuracy compared to (OPNET, NS-3, and NS-2 or even MATLAB) simulators [11–13]. QualNet simulator is a high-fidelity network analysis program that predicts the efficiency of wireless, wired, and mixed-platform networks and networking systems. QualNet supports thousands of network nodes to be simulated. QualNet provides unrivaled portability of the software and versatility of interfaces. And also QualNet operates on UNIX, Windows, Mac OS X, and Linux parallel and sequential operating systems [14, 15].

Throughout the calling growth, VoIP technology will occupy a huge internet communications bandwidth, resulting in a large increase in demand for the network's transmission capacity. In current mobile network pricing schemes, where billing relies on usage frequency, a considerable rise in network resources, would quickly lead to a dramatic increase in cell network rates. A major VoIP technology challenge is the trade-offs between the volume of data flow and the level of touch in voice communications [3]. Our main problem how to pick the best signaling protocol that achieves high efficiency over wireless networks to accommodate the

applicable call density, by taking into account the different metrics to be measured which is Session total Bytes, Establishment time, Sent, total Bytes received RTP average end to end delay and Overall Throughput. The paper achieves the flowing points.

- Factors that determine voice quality include the choice of codec, echo control, packet loss, delay, overall throughput, and the network's design of the network Packet loss causes voice clipping and skips.
- The Quality of Service (QoS) can be achieved by managing router queues and by routing traffic around congested parts of the network.
- Improve efficiency by identifying the best VoIP signaling protocol that achieves high performance over wireless networks.

The rest of the paper is organized as following the second section, focusing on related works. The third section discusses the methodology. The fourth section simulation scenario and the fifth section presents the result and analysis, and the last section sixth concludes the paper.

2 Related Works

Many studies were performed to evaluate the performance of VoIP by changing the codecs other studies discussed the signaling protocols of the VoIP.

Like in [16] study conduct on the performance of VoIP service with the G.729 and G.711 codecs based on the same network topology focus only on the QoS of the Voice over IP platform and the study conducted without taking into account signaling protocols and limitations of the simulation software plus the outcome.

The researchers in [9] evaluate (VoIP) performance and two signaling protocols SIP and H.323, the study performed in a homogeneous 802.11e, and the findings showed the SIP protocol's efficacy in terms of call setup time, but the study focused on MIPv6 and its limitations, especially on mobility speed, and no overall throughput.

The results in [3] concluded that: SIP voice quality is higher than Red5 through conducted comparison of SIP and Red5 flow, and on voice quality acceptance level of the general public In addition to SIP, the focus was put on a different protocol were not H.323.

As discussed in [17] paper draws a comparison using an (OPNET Modeler) between four combinations of H.323 and SIP Protocols and with G 0.711 and G.729 codec, The combination H.323 and G.729 proved to be the most appropriate one in the connection process, the study limitation was the protocols were not individually performed.

A study performed in [18] to compare the performance in the network of two types of codecs G.711 and G.729 on IEEE 802.11ah standard for voice over IP services, the evaluation concentrated on the following scales packet, delay ratio,

average delay, and throughput, the study conducted using network simulators (NS-3), hence the research is a little out of scope.

3 Methodology

To create and manage multimedia sessions, signaling protocols are implemented. There are currently two standard protocols on the market that are commonly used, SIP and H.323 [19]. Our research went through the behavior of two VoIP protocols modulated, and the performance will be evaluated through building scenarios on a QualNet simulator, using the Random Waypoint Mobility Model (RWP) the random nodes will choose the direction and its constant speed in each movement randomly and then after shortstop period it will choose another destination and another constant speed, therefore, it considers close to reality and better than choose various steady scenarios [20]. To investigate the performance of this protocol mechanism, the statistical results will be analyzed to come up with a comparison between the protocols, as in Fig. 1.

4 Simulation Scenario

In the simulation process, the experiment scenarios implemented using the simulator QualNet by following steps and as illustrated in Fig. 2.

- Firstly configure the general properties which apply to the whole Scenario.
- Specify the network topology by creating subnets, placing nodes, and define g nodes mobility
- Configure the protocol stack for individual nodes or groups of nodes as necessary.

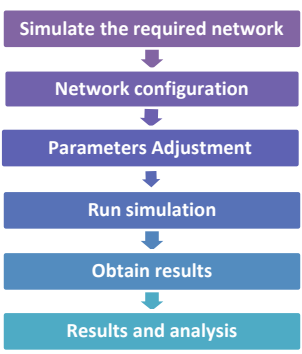


Fig. 1 Methodology steps

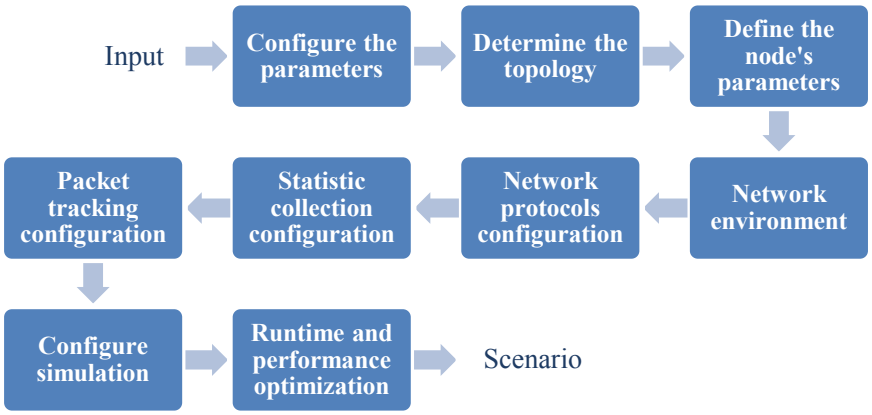


Fig. 2 Creating a scenario

- Configure parameters for collecting simulation results and controlling runtime performance.
- Analyzing the results.

The last step is to use the file editor to compare the performance of routing protocols based on the performance metrics selected.

Our Scenario in Fig. 3 and general parameters, were considered a WLAN network that has 10 nodes. Three of them are RWP mobility nodes. Node 10 is used to perform a generalized function for both of the VOIP networks (H.323 and SIP). Node 1, 2, 3, 4, 5, and node 6, 7, 8, 9 are connected to wireless network1 and wireless network2, respectively. Node 10 is locally connected to nodes 5 and 6.

Fig. 3 Wireless ten nodes with VOIP application scenario

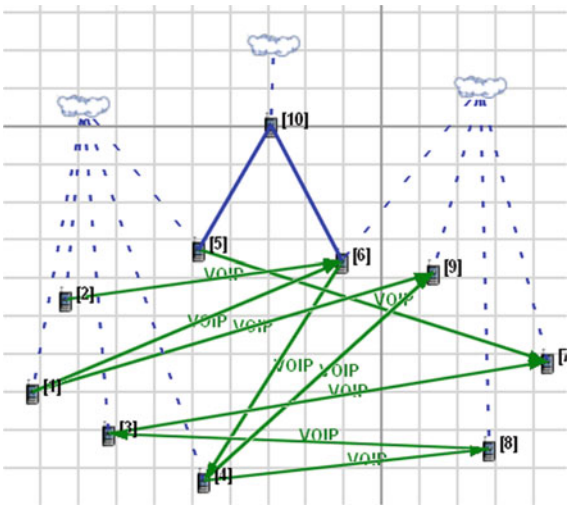


Table 1 Scenario parameters

Parameter	Specification
Number of nodes	*10*
Simulation area (meters)	*1500 × 1500 m*
Mobility	*Nodes(1,7,8) random, others with no mobility*
Application	*VOIP with SIP and H.323*
Codec	*G.711*
Simulation time	*For both protocols 256 s* *Total 134 s*
Network type	*Two wireless networks*
Mobility type	*RWP*

We have chosen the mobile nodes as hosts 1, 7, and 8. Host 7 and 8 have a slow movement in comparison to 1. Then we simulated the scenario in the QualNet simulator two times for different protocols (SIP and H.323), the scenario parameters are listed in Table 1.

5 Result and Analysis

Experiments for this study are to analyze SIP and H.323 performance. The purpose is to identify the best transmission protocol. From our experiments, As shown in Figs. 4 and 5. The collected result analyzed and used to compare the performance. The scenario was simulated in the Qualnet environment two times the total simulation time has taken for both protocols 256 s with RWP mobility type. The initiator session establishment time for the SIP application is 125 s, whereas H.323 has required 131 s to establish the session.

The receiver session establishment time for SIP and H.323 application is the same as the initiator session establishment time. As we notice in the initiator process of SIP has sent more Bytes 60% compare to H.323 for the same period.

Correspondingly, the receiver packets were the total bytes sent increased by 53% in SIP. while dropdown around 59% in H.323.

This comparison is based on a G.711 codec. depicts the delay amount in SIP protocol was 0.05 s. The following equation is used to calculate the average end-to-end delay, where

$$T_{E2E} = (T_R - T_S) \tag{1}$$

But the delay period is not fixed In VOIP applications it is different for the different codec.

Finally, the overall throughput in Fig. 6, of the two signaling protocols concerning the duration of audio length. The result gives a good comparison between

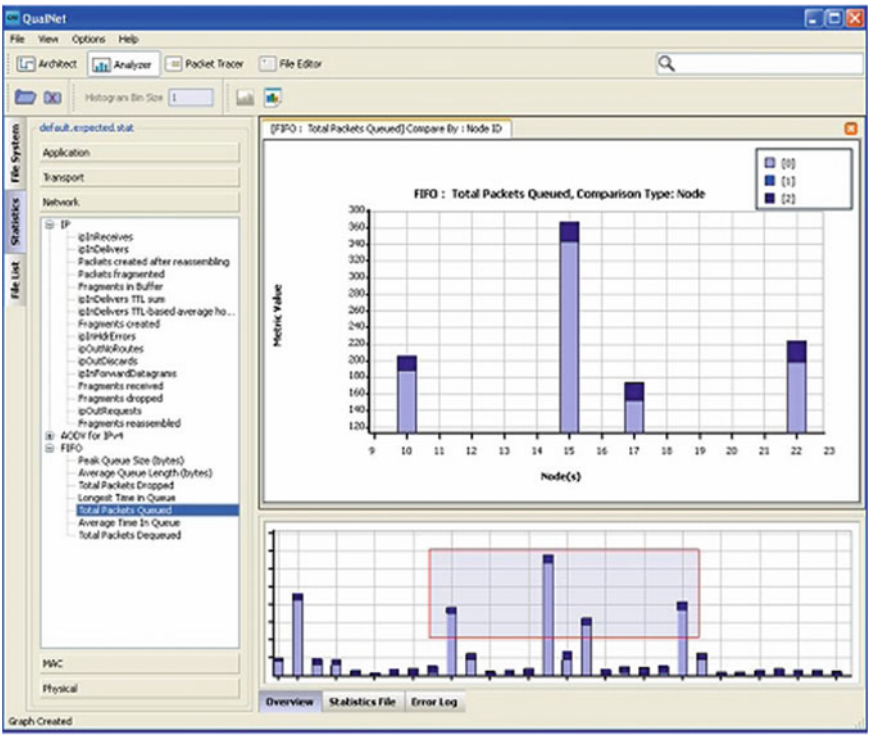
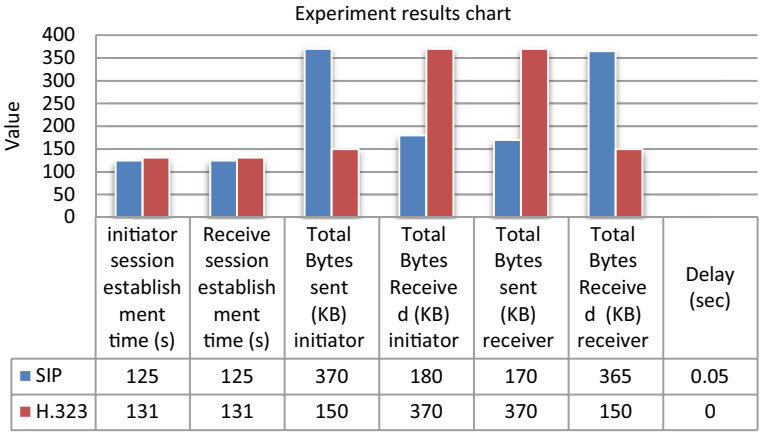


Fig. 4 QualNet analyzer



	initiator session establishment time (s)	Receive session establishment time (s)	Total Bytes sent (KB) initiator	Total Bytes Received (KB) initiator	Total Bytes sent (KB) receiver	Total Bytes Received (KB) receiver	Delay (sec)
SIP	125	125	370	180	170	365	0.05
H.323	131	131	150	370	370	150	0

Fig. 5 Experiment results

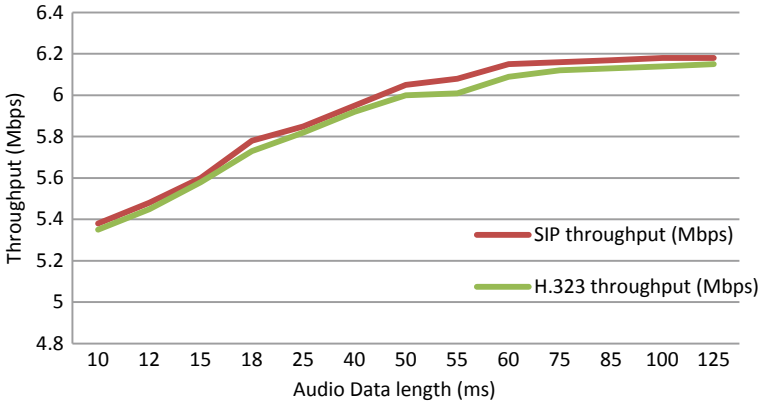


Fig. 6 The overall throughput

the protocols for the matter of throughput calculation. In SIP the maximum throughput was 6.182 Mbps at the duration of 123 ms and 6.122 Mbps at the same duration for SIP and H.323.While SIP shows absolute superiority all duration.

6 Conclusion

In this paper, the two common signaling protocols SIP and H.323 simulated by QualNet-based network to evaluated and analyzed both protocols performances in VoIP and 802.11a and wireless network environment, Based on the proposed simulation and overall SIP superiority in throughput especially in long period audio session, moreover the establishment time and session initiation are better in SIP, With an unnoticeable delay in the end to end delay result which would not impact the protocol performance. The future direction includes more different signaling protocols for comparison.

References

1. Islam SKH, Vijayakumar P, Bhuiyan MZA, Amin R, Rajeev VM, Balusamy B (2018) A provably secure three-factor session initiation protocol for multimedia big data communications. IEEE Internet Things J 5(5):3408–3418. https://doi.org/10.1109/JIOT.2017.2739921
2. Nikooghadam M, Amintoosi H (2020) A secure and robust elliptic curve cryptography-based mutual authentication scheme for session initiation protocol. Secur Priv 3(1). https://doi.org/10.1002/spy2.92
3. Lin CN, Lin TL, Chen TS, Chen J, Chen WJ (2016) VoIP communication quality and flow volume preference—a SIP and Red5 example. In: 2016 3rd international conference on systems and informatics (ICSAI), Dec 2016, pp 782–786. https://doi.org/10.1109/ICSAI.2016.7811057

4. Liu H, Mouchtaris P (2000) Voice over IP signaling: H.323 and beyond. IEEE Commun Mag 7. http://files/1126/Liu and Mouchtaris-2000-VoiceoverIPSignalingH.323andBeyond.pdf
5. Goode B (2002) Voice over internet protocol (VoIP). Proc IEEE 90(9):1495–1517. https://doi.org/10.1109/JPROC.2002.802005
6. Hamdi MM, Mustafa AS, Mahd HF, Abood MS, Kumar C, Al-Shareeda MA (2020) Performance analysis of QoS in MANET based on IEEE 802.11b. In: 2020 IEEE international conference on innovation technologies, INOCON 2020, no November, 2020. https://doi.org/10.1109/INOCON50539.2020.9298362
7. Aboode (2016) A quiz management system based on P2P near-field communication on android platform for, pp 14–16
8. Nikoukar A, Hwang I-S, Liem AT, Lin Y-H (2015) Local-aware H.323-based VoIP service in EPON. In: 2015 international conference on electrical engineering and informatics (ICEEI), Dec 2015, pp 661–664. https://doi.org/10.1109/ICEEI.2015.7352581
9. Khiat A, El Khaili M, Bakkoury J, Bahnasse A (2017) Study and evaluation of voice over IP signaling protocols performances on MIPv6 protocol in mobile 802.11 network: SIP and H.323. In: 2017 international symposium on networks, computers and communications (ISNCC), Dec 2017, pp 1–8. https://doi.org/10.1109/ISNCC.2017.8072037
10. Gao X, Gong P, Liu Y, Meng Q, Li C, Shan D (2019) An interactive interface of QualNet for DigSilent in power grid simulation. In: 2019 21st international conference on advanced communication technology (ICACT), Dec 2019, pp 330–333. https://doi.org/10.23919/ICACT.2019.8702045
11. Hamdi E (2018) BER Vs Eb/N0 BPSK modulation over different types of channel. Aust J Basic Appl Sci 12(5):31–38. https://doi.org/10.22587/ajbas.2018.12.5.7
12. Khan S (2017) Asterisk server and OPNET platform, pp 28–31
13. Prajwala J, Mathew R, Taj N (2018) Analysis of VoIP traffic over LTE for different Codecs. In: 2018 3rd IEEE international conference recent trends electronic informatics communication technologies RTEICT 2018—Proc, pp 1858–1862. https://doi.org/10.1109/RTEICT42901.2018.9012408
14. Aireen G, Mohan CE, Pooja CH, Pooja FT, Raghuram KM (2017) Wireless network simulation and analysis using qualnet. In: 2017 2nd international conference on communication and electronics systems (ICCES), Dec 2017, pp 251–255. https://doi.org/10.1109/CESYS.2017.8321276
15. Abood MS, Thajeel IK, Alsaedi EM, Hamdi MM, Mustafa AS, Rashid SA (2020) Fuzzy logic controller to control the position of a mobile robot that follows a track on the floor. In: 4th international symposium on multidisciplinary studies and innovative technologies, ISMSIT 2020—proceedings. https://doi.org/10.1109/ISMSIT50672.2020.9254417
16. Sombie AT, Poda P (2018) Towards enhancing the performance of a VoIP infrastructure for limited-resource structures. In: 2018 1st international conference on smart cities and communities (SCCIC), Dec 2018, pp 1–4. https://doi.org/10.1109/SCCIC.2018.8632842
17. Rattal S, Badri A, Moughit M (2013) Performance analysis of hybrid codecs G. 711 and G. 729 over signaling protocols H. 323 and SIP. Int J Comput Appl 72(3):30–33. https://doi.org/10.5120/12476-8871
18. Nur'afifah RS, Perdana D, Istikmal (2019) Comparative analysis of codec G.729 and G.711 on IEEE 802.11AH with MCS and raw slot change mechanism for VOIP service. In: 2019 4th international conferene on information technology information system and electrical engineering ICITISEE 2019, pp 379–384. https://doi.org/10.1109/ICITISEE48480.2019.9003940
19. Naeem MM, Hussain I, Saad Missen MM (2020) A survey on registration hijacking attack consequences and protection for session initiation protocol (SIP). Comput Netw 175:107250. https://doi.org/10.1016/j.comnet.2020.107250
20. Bettstetter C, Hartenstein H, Pérez-Costa X (2002) Stochastic properties of the random waypoint mobility model, p 7. https://doi.org/10.1145/570758.570761

Bringing Blockchain Technology in Innovating Industries: A Systematic Review

Daimler Benz, Muzaffar Hamzah, Mohd Fahmi Ghazali, and Mohammad Fadhli Asli

Abstract The mainstream digital technology uses significantly changes how industry protect their records due to new security and vulnerability challenges. Leveraging cryptography, blockchain technology enables large agile value chains, transaction transparency, and flexible integration with many digital technologies like Internet-of-Things and cloud computing. The capabilities of blockchain technology further empower secure transferable wide-range collaboration between industry allies and public sectors. Recognized by its potentials in driving innovation in industries, blockchain is expected to play significant roles in industrial transformation. This paper reviews the significant developments of blockchain applications in information systems of many industries. The review aims to provide insights on the advances of blockchain industrial applications including use cases, classifications, and emerging trends. We conducted a systematic review on published blockchain works from high-quality sources using formulated review questions, selection criteria, and processes. Summarizations of blockchain applications across many industries are presented including use cases, objectives, and application examples. Moreover, we classified and characterized the innovative features of presented blockchain applications based on the application's distinct characteristics. Subsequently, summarization of emerging trends in blockchain applications based on recent industry use cases and implemented features are discussed. This paper highlights the consequential advances of blockchain in industries that eventually lead to research opportunities for future undertaking.

Keywords Blockchain · Industry 4.0 · Distributed Ledger · Use cases

D. Benz · M. Hamzah (✉) · M. F. Asli
Faculty of Computing and Informatics, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia
e-mail: muzaffar@ums.edu.my

M. F. Asli
e-mail: fadhli.asli@ums.edu.my

M. F. Ghazali
Faculty of Economics and Management, The National University of Malaysia, Bangi, Malaysia
e-mail: fahmi@ukm.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_33

1 Introduction

As we move into the new era of technological advances, most of our daily activities are digitalized or mechanized for better convenience. This majorly differs from previous advances from our predecessors in the 90s to early 2000, where most of our daily tasks are carried out manually. These tasks include monitoring, recording, managing, organizing and other repetitive tasks such as buying a ticket to watch a movie or to watch a baseball game. Recent literature show that many systems have been implemented in a digital format, from the automated toll system to the attendance system implemented in the workplace and higher learning institutions [1, 2]. Blockchain technology gained attraction in the light of rising popularity of cryptocurrency. Over the years, blockchain has gained popularity due to popularity of cryptocurrency. The application of blockchain has developed since then and has been incorporated in several real business cases within industries. Current blockchain applications are observed in the financial sector, food supply chain, international documentation, and music distribution.

Blockchain technology can bring forth tremendous innovation in industrial transformation to industry 4.0 [3]. There is yet to be a systematic review on the trend of blockchain development in various industry enclose with use cases and classification. And by the year goes, new use cases and emerging industry is expected as their implementation scale is large and need to be observe so that blockchain matters can be understood in depth to enhance the mainstream adoption of its technological environment.

The aim of this paper is to report the overview of significant developments in blockchain usage within information systems in various industrial sectors. Insights gained via this overview help profound understanding about the blockchain development and identify innovative factors that blockchain brought to the industries. In this study, we carried out a systematic review of recent blockchain literature to gain insights on industrial innovations brought by blockchain technology development. First, we examined each industry and compose it throughout relevant blockchain use cases. Next, we observed and classified blockchain features that were set as distinctive characteristics in these industries. Besides, this review examined the recent trends on blockchain use cases and blockchain features that are frequently used in industrial works. To summarize, the contributions of this paper are:

1. Summarization of overviews on blockchain technology use cases including type of use cases, development context, and tangible benefit of blockchain for each use cases at any form of innovation in many industries.
2. Classifications of blockchain concatenation according to recent industry operational. For each of the classification, we discuss the performance that generally has provide a solution onto industry flaw.
3. Summarization of emerging trends covering blockchain use cases and its classification.

The remainder of this paper is organized as follows. Section 2 introduces the underlying concept and background works of blockchain implementation. Section 3 describes the methods used in conducting this systematic review including research questions, review workflow, and selection process. Section 4 then presents the results of our systematic review based on each formulated research question. Section 5 then provides detailed discussions and summarizations of findings based on review results. Finally, the conclusion of this review paper is presented in Sect. 6.

2 Background

Securing communications and transactions is a crucial aspect of information security especially with the mainstream adaptation of digital information technologies. Leveraging cryptography, the concept of blockchain was introduced to help to secure securing public transaction records with manipulative-resistant design. Blockchain is related to a database technology called distributed ledger that catalog a digital event record and interlinked by each data collection on every device that integrates with the transactions over the database. This data is protected as private property in blockchain, no computing or person and even the admin can change this protected data after the approval transaction takes place for the data to be recorded [4]. Blockchain essentially provides a public network for everyone to use by adding their data into the system, besides there are also includes private blockchain to be applied [5]. The data within the chain itself use 'distributed concurrence' which the data cannot be stored into the blockchain database if and only the agreement has been verified, whether a new data is inserted into specific blocks that are placed according to blockchain transactions and it is lawful or illegitimate to be recorded to the chain. Nowadays, massive amount of data is interconnected globally, and it is a valuable asset in various industry to be protected [6]. Researcher in academia and industry realize the benefits of blockchain through strengthened security, better traceability, and accountability. For example, an investigation on integrating blockchain technology and supply chain discovered that the food industry needs blockchain to ease their customer for checking the origin of product details [7]. Another study found that the educational system can be integrated with blockchain to record a permanent credential for students' achievement in the academia area [8].

We found 6 studies that reviewed the leading edge of the blockchain in various industries. One of the papers is focused on blockchain application and use cases that affect blockchain design, while another is focused on blockchain implementation for various industry applications. One paper discussed the emerging management of supply chain, logistics, and transportation using blockchain application, another one papers elaborate the vital factors on blockchain implementation in supply chain

management (SCM). Two papers focused on the blockchain aspect, which is industry 4.0 in blockchain and blockchain on security and privacy. The paper mentioned above is divided into four categories like use cases, industry application, industry 4.0, security, and supply chain as lists in Table 1.

A survey [9] reviewed the influencing factors brought from blockchain application and lay claim that the features of blockchain application offer endless new, while its use cases bring a solution for real-world problems. Meanwhile, a survey [10] highlighted the obtainable opportunity, benefits, and challenges in various industries for utilizing blockchain. The emerging of blockchain applications in industry and services (e.g. supply chain, logistics, and transportation) is discussed in a study [13], by leveraging the framework of technology, traceability, trust, and trade. Another review [14] elaborated the vital factors towards industry and put into words that the factors found in their study could be applied for SCM development, indicating that latest technology can be implemented to enhance performance, lowering costs, improve safety, and strengthen productivity by adopting blockchain in SCM. A study [11] focused on the capability and solution of blockchain when integrating with industry 4.0 based applications, and they found blockchain solved much current security that failed to cope with traceability, network inactivity, information processing, and more. Besides, a systematic review [12] presents summarizations of the aspect of security and cyber-attack within the blockchain, and their analysis results indicate blockchain is under imminent threat. These surveys and reviews discussed many of the related aspects of our review focus. Nevertheless, these papers did not discuss the blockchain leading edge based on trends within use cases and its utilization in various industries.

The premise offered by blockchain technology is the elimination of unnecessary third parties. Blockchain significantly reduces transaction fees and time, clearing, and settlement systems. Blockchain turns every transaction to be open, direct, and convenient, thus re-establishing trust among purchasers and providers. Moreover, blockchain features the elements of security, trustworthiness, and transparency. Awareness in making collaboration between industry allies, competitors, and customers contribute to the quality of values, including capital, intellectual, and technology investment. The transparency and security of blockchain, or distributed ledger technology, lower costs for many programs and projects. These instances do occur for countries with weak or underdeveloped legacies of telecommunication and financial infrastructures.

Table 1 Categorization of existing reviews on blockchain implementation

	Use cases	Industry application	Industry 4.0	Security	Supply chain
Survey	[9]	[10]			
SLR			[11]	[12]	[13, 14]

3 Methodology

In this section, we describe our methodology in conducting the systematic review including research questions, review process, eligibility criteria, search strategy, and data collection. Referring to PRISMA [15] and reference models [16, 17], we conducted a systematic review to deliver an overview of the blockchain development. Additionally, we refer to the existing model concept [18] for conducting this study, which consists of the review process, the following paragraphs and sections process illustrated in Fig. 1.

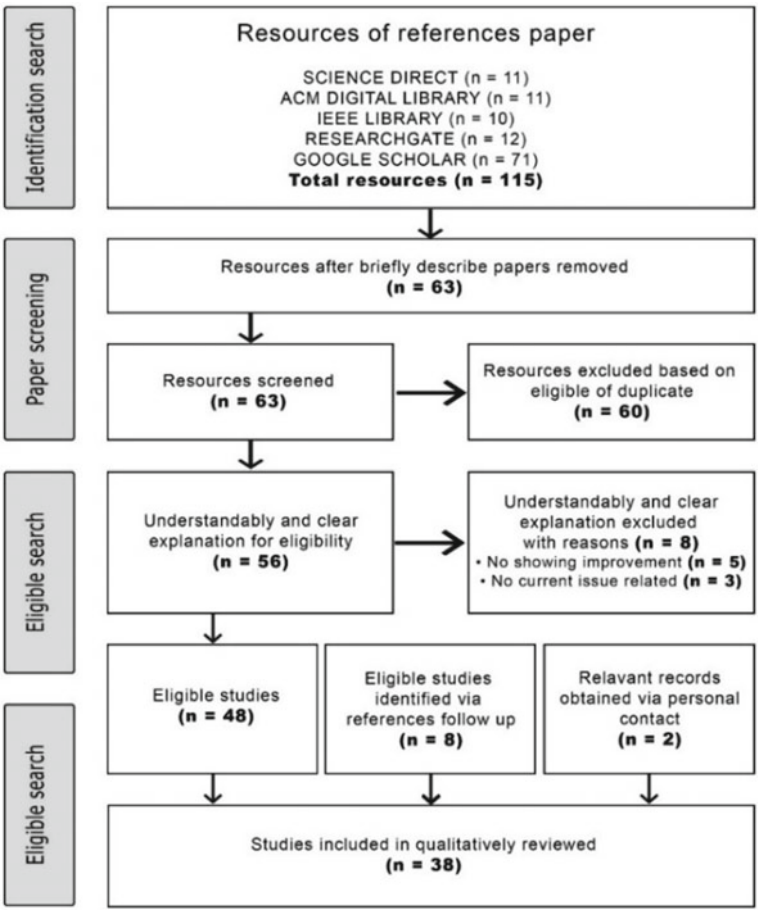


Fig. 1 Review workflow of this systematic review

3.1 Review's Research Question

A systematic review begins with the formulation of guiding research questions [16]. Therefore, the following research questions by this study are as follows:

RQ1: What are the recent developments of blockchain that innovate productivity in various industries?

RQ2: What are the classifications of improvement brought by blockchain technology in industries?

RQ3: What are the emerging trends of blockchain utilization in industries?

RQ1 aims to identify the classification of blockchain adoption and operations in the various industrial sector. Although not all the industrial sector implemented blockchain, there are several industries still uses traditional transactions. Hence, it is useful for classifying blockchain capability relation to all industries, include the early stage to get used blockchain. By highlighting the blockchain classification might help to inform the improvement directions that blockchain brings in various industries, as well to increase blockchain adoption.

RQ2 aims to gather the existing development brought from blockchain features for selected industrial sector. Classifying the outcome of use cases enable other researchers and experts to interpret for a better understanding on the main research area, that could contribute towards more investigation on blockchain use cases in future research.

RQ3 aims to examine recent works to offer insights on the recent trends of blockchain. New technology, use cases, and development might be introduced in the future. Determining recent trends of blockchain uses by various industries eventually supports other researchers and experts to obtain insights on blockchain development over the half decades.

3.2 Review Protocol and Phases

This analysis applies the developed standard guidelines [19] to assess planning for review, to analyzing the review by interpreting selected studies through e-databases, and to outcome the final review. This protocol process is illustrated in Fig. 1.

3.3 Inclusion and Exclusion Criteria

This review selected peer-reviewed papers that fulfill the predefined eligibility criteria to be included in the review. The descriptions and justifications of the selection criteria are shown in Table 2.

Table 2 Inclusion and exclusion criteria in review selection

Criteria	Justification
IC-1: The paper describes the use of blockchain in different industries	This review investigates relevant studies on blockchain adoption in various industry
IC-2: The paper clearly illustrates the improvements that have emerged due to blockchain in various industries	The aim of this paper is to deliver an overview on significant developments of blockchain technology industrial sector
EC-1: Papers describe the implementation of blockchain in today's technology	Paper that only overviewing blockchain implementation in general do not always contains an integration with industry aspect
EC-2: Duplicates or identical research regarding the usage of blockchain and their advantages to the various industries	Duplicate content is always subject to confidentiality and uncertainty, thus we compared date published and status of the author

3.4 Information Search Strategy and Sources

The papers were extracted from online electronic databases as demonstrated in Table 3. Two main search keywords were used: "blockchain" and "industry". Since not all papers have similar topics, the related papers in which the point are relevant to the topic were manually searched to identify any additional relevant work to be included in the review.

4 Results

This section presents the results of data collection with regards to each formulated research questions.

Table 3 Distribution of papers by sources throughout screening process

Sources	Identification search	Paper screening	Eligible search	Selected for review
ScienceDirect	11	10	9	9
ACM DL	11	10	9	9
IEEE Xplore	10	8	8	7
ResearchGate	12	10	8	6
Google Scholar	71	37	35	7
Total	115	63	48	38

4.1 Results for RQ1

What existing development of blockchain innovates productivity in various industries? Out of 38 studies, two studies [20, 21] are discussing blockchain-based use cases practical in the emerging industry, followed by the use cases of supply chain [22–24], and financial [25–27] as well music industry [28–30] with three studies each. One study is focusing the document management [31] while another gaming industries [32], and another one is conducting humanity resources management [33]. With some of the studies we got above, we analyzed it together with an innovation-based paper [34].

The application of blockchain allows for a decentralized and distributed environment that does not require central authorities [4, 35]. Most of the blockchain application is associated to data management and authentication [36]. All data transactions are trustworthy and secure because of cryptographic principles in the blockchain [5, 37]. Nowadays, blockchain technology has penetrated different domains gaining recognition for wider audiences leading to an increase in the proposed services and software application [38]. These industries are labeled based on their category sectors as shown in Table 4.

(a) Supply Chain

Problem and challenges might happen at anytime and this leads to postponement activity in supply chain management. Universal and global practicalities in the

Table 4 Classifications of blockchain use cases

Use cases	Usage of blockchain	Industrial sector
Supply chain	Traceability for products Quality assurance	• Walmart: Tracking food safety and supply chain • DNV GL MyStory: Track provenance of wine from grape to bottle
Financial	Microloans International trade Retail lease guarantees	• Twiga Foods: Small Business Microloans in Africa • ANZ and Westpac: Retail lease guarantees • HSBC: Improve efficiencies and reduce time
Document Management	Document distribution Digital documentation Digital certificate	AB InBev, Accenture, APL, Kuehne + Nagel: Eliminate printed shipping documents • Accedible: Digital educational certificates • North and South Korea: Recording the Peace Declaration
Music Streaming	Music distribution Track and store digital. information	• Spotify: Validate and register transaction without a need for a central authority • Musicoin: Streaming catalog of music from independent musicians
Humanity Welfare	Welfare distribution	• UN World Food Programme: Refugee relief
Gaming	Digital collectible assets	• CryptoKitties: Digital collectibles assets

context of supply chains are incompatible, resulting difficulties in monitoring or organizing manpower and technology. Moreover, human workers are prone to technical mistakes either intentionally or unintentionally on global logistics service. Blockchain technology transforms the workflows for logistics in the supply chain to be more efficient and optimized. A study [39] explained that this optimization is designed to facilitate accurate inventory tracking, which allowing for further service licensing, improves accountability from source to production point of consumer satisfaction.

As the food supply chain management is important in addressing current limitations on the blockchain, a recent work [40] investigated the effect of blockchain on food supply and found that blockchain is used by Walmart and DNV GL as a digital assurance solution by providing traceability to its products mainly for food supply in their supply chain. In addition, a study [41] critically examined the supply chain management relationships and stated that blockchain featuring traceability that can be integrated to every secure transaction entity in the domain supply chain. The transaction is validated and provided in the blockchain by immutability functions, only mutually accepted participants are involved in it [42]. Overall, the adoption of blockchain to supply chain or logistic process strengthens the performance, reduces cost, augments risk measures, and improves productivity dramatically [14].

(b) Financial

From the financial perspective, some aspects can be achieved by using blockchain such as linking banks around the world without middlemen. A recent work pointed that financial services are struggling to cope with outdated and obsolete financial operations, and also gaps in security systems slow payment transactions and limited transparency [43]. These financial problems can be solved using blockchain to transform significant financial markets into operations with faster accountability processes, transparent technology [44], as well as efficient and secure industrial systems and infrastructure systems [45].

Blockchain is used by Twiga Foods to enable Small Business Microloans in Africa which has growing financial services to overcome the common issues in least developed countries. Microfinance is one of the critical service tools that stand against poverty as an important behalf of IFC's to make a quality transformation of a vibrant private sector in the country. Within the microloans itself, it is focused on the peer-to-peer approach. Peer-to-Peer (P2P) lending, knowing as "social lending," the functionality is to make transactions among individual-to-individual to lend money.

HSBC bank also uses blockchain for international trade with Corda or known as R3. A study [46] claimed that Corda is third wave blockchain which promises security and efficiency, especially in the way it operates. By using Corda, an issue letter of credit can be made especially for US for an agreement of food, agriculture film, and Cargill. Since it normally takes the range of 10 days to complete the agreement, blockchain reduces the time for consensus, which takes only 24 h to complete the agreement.

ANZ and Westpac use blockchain technology to retail lease guarantees because it is an open source, secure in transmitting information, and esteem. Blockchain able to make the banks to be more productive, where the expenses of instalment be prepared immediately, produce new items and administrations aware of producing new income streams. Banks were using blockchain to identify each individual and their authorization as it plays a vital part both in the banking business and the customers [47].

(c) Document Management

A study [48] examined the digital learning credential assessment and management on blockchain and found that the everyday activity in documents management are processed either by people or automation machine such as digital certificate digital learning credential. Also, a study [49] investigated the applicability of blockchain-based electronic transfer bills and found a long heterogeneous alignments document like bills sometimes are difficult to index in a current traditional database. A recent work [50] explained the emerging research and opportunity of smart contract management in which document management is one of the blockchain most incontestable features, integrating and implementing these technologies not only has improved document management but it also enabled standard audibility as most other of its activities are provable and data distribution. To address the data distribution issues, decentralized blockchain features for metadata are suggested [51] that supports the key features and examine consequences for digital documents management and information conservation.

Blockchain is used by AB InBev, Accenture, APL and Kuehne + Nagel to eliminate printed shipping documents and transition to digital copy. Throughout this research of documentation review processes, AB InBev, Accenture, APL and Kuehne + Nagel determine and re-allocate of accountability, forecast risk and ownership information with blockchain technology trust. A recent study [52] discussed the blockchain capabilities in which able to make the process management faster for the entire flow of transport documents, making data entry requirements more productive. Moreover, blockchain capabilities make easy for data amendments during shipping, align monitor and overcome unpredictable cargo processes of penalties conflict or cases complaint from most customer that are not satisfying for the services [53].

Blockchain is also used for international documents—North and South Korea use blockchain to record the peace treaty. International documents are records pertaining government policies, thus extremely crucial to secure via safe platforms to ensure longevity and integrity of the documents. A smart contract is one of the blockchain applications that execute as a program by consensus protocol [54, 55]. And there are many blockchain technologies that can store smart contracts such as Stellar, EOS, Cardano, NEO, Tron, and Ethereum. Smart contracts are recorded on their platform to have some form of transparency, security, and no possibilities of bias, manipulation, and error [56, 57].

(d) Music Streaming

The digital world has revamped the music industry, and the implementation of streaming platforms has accelerated the speed of the value chain. However digital duplication, fake cloning, infringement of property rights of production companies is causing disruptions that causes losses in revenue production. A study found that blockchain-based media corporations enhance the effectiveness of rights, which offer an innovative platform to redeem attraction to the ecosystem [58]. One of the biggest issues throughout the music industry today is the inexistence of worldwide registry of creative music. As blockchain developing to be more mature, it facilitates the entities that are involved in making or engaging with music with a faster and more streamlined system. Blockchain can transform both the database as well as the network infrastructure, thus reduces the issue of intellectual property piracy, property rights dispute.

Blockchain is used by Spotify to unlock hardware limitations for music streaming. Besides, a study [59] investigated the impact of blockchain in the music industry and found that the uptime is also almost 100%, the transfer rate is the highest, data integrity is always confirmed, service availability is now for everyone, the storage size is almost unlimited, and all the limitations are enabled to achieve freedom transaction from a centralized authority. With the rise of blockchain, latency is decreasing significantly, uptime 100% is achievable and high transfer rate data and transaction is possible. Musicoin uses blockchain for streaming catalogs of music from independent musicians by tracking and saving analytics broadcast, with distribution to pay the artist in real-time. Blockchain allowing its distribution database to record digital information without enabling the modification mode toward its goal. Musicoin uses blockchain to power a decentralized peer-to-peer platform, hosting music available to everyone and transactions become transparent and secure.

(e) Humanity Welfare

Blockchain is useful to the UN for its world food program to refugee relief, a study [60] reported that blockchain application can expand outside of business and government use cases and be applied to the social realm of improving the lives of individuals. The blockchain laboratory posts that 15 UN agencies undertake the blockchain initiative. Several campaigns have chosen to use blockchain technology for various humanitarian disasters, especially to the crisis of refugees. For example, in a recent review [61], the blockchain initiative program aims at over 100,000 refugees in Jordan living with their own camp known as the Azraq refugee camp. Moreover, most assume that these refugee issues can be resolved by an approach to self-sovereign identity (SSI). This method seeks to assign control to individuals by enabling them to register their own identity data. This registration can be performed appropriately with blockchain [62], that SSI platform may also permit refugees that unable to acquire government identification to create a composite ID through multiple assertions.

(f) Gaming

A study [63] found that blockchain technology have a great impact on digital games, particularly people with really large multi-player scenarios. Additionally, over recent years the game industry has expanded rapidly, mainly because the computer game world has managed to keep up and is attempting to address the new technical advances [64]. However, traditional game technologies are implemented in enclosed loops, preventing external creation of another version and access for the core concept or use the current data (domain). A study [65] claimed that the most core concept and current data (domain) are managed by smart agreements, which may be extracted and replicated. Blockchain game domain is circulated between 'players' rather than controlled within a central server. Blockchain games either generate by using their own token or a current token [66]. Additionally, blockchain is used by CryptoKitties as digital collectible assets. Gamifying features like CryptoKitties leverage on blockchain's unique applications. A crypto collectible is a digital piece of non-fungible data that is specific to cryptography. Therefore, the overall tokens in circulation are the same when it comes to a regular cryptocurrency [67]. While CryptoKitties itself is not a digital currency, yet it does offer the same security. CryptoKitties is an online game built over the Ethereum platform where players buy images of cats with specific traits and 'breed' them to produce new cats that can have rare traits.

4.2 Results for RQ2

What classification of improvement is brought by blockchain technology in various industries? Some industries are hesitant in transforming their conventional transactions practice due limited blockchain knowledge and risks consideration. The classifications of blockchain features are explained as shown in Table 5.

Table 5 Classifications of blockchain features

Classification	Definition
Security	Every one of the blocks of information, for example block or hash (#), are secured and bound to one another by utilizing the cryptographic standards
Trust	Blockchain overcomes the issue of trust by eliminating human intervention from the equation
Traceability	The inviolability of blockchain gives traceability chain an even more convincing value
Efficiency	Efficiency concerns the level of performance of using blockchain, particularly how efficient blockchain application is in bitcoin if used based on time and energy. Fundamentally, blockchain reduces the number of wasted resources used to produce a given number of goods or services
Reliability	The hallmark of the blockchain system is reliability, which concerns how much the data are retrievable when there is still somehow connection that exists somewhere. Blockchain secures the data and allows us to access even if there is a problem occurring in network mesh

5 of the studies focus on the capability of blockchain security [68–72]. Besides 4 studies classification focusing the blockchain traceability in the supply chain [73–75] and its issue challenges along with solution [76, 77]. Besides, another 3 studies are discussing blockchain trust [78–80], and as well efficiency [81–84]. Another 2 studies are discussing reliability for blockchain [85, 86].

(a) Secure

Blockchain functions as a distributed database where all information is stored through each partition chain named "Hash" or "Node". Also, it is called a "Decentralized System" [4] where there is no central administration, no modification to existing data, and no deletion of data. The information can be retrieved by the user to display or show a record. In [87] interpret that there is no secrecy in the blockchain implementation, and it is called the "Open Ledger System". Unlike centralized systems, such as banks, the system is controlled by them that is certainly easy to manipulate. Therefore, it is extremely crucial to secure the documents on a safe platform by using Ethereum [88], which is proven to be secure yet to allow developers to develop decentralized applications and deploy them. Ethereum is an application that utilizes blockchain security, it has the capability to maintain and storing transaction, so the apps have no authority to steal data. There are some vulnerabilities in the existing Ethereum [55] with 8833 out of 19,366 and this may lead to losses. Besides, a study [89] highlighted that it promises a protected centralized platform to make things easier with all consumers and third parties to communicate, exchange, and sharing information.

(b) Trust

A study [90] reported that blockchain incorporates 'trustworthiness' through a distributed computer network that seeks to achieve a common set of goals, by eliminating third parties, reduce costs and generate more profit for various industries. Furthermore, trust in blockchain becomes a central factor and important to emerging technology [91], many researchers require thorough research to verify the heuristic relation in both trust feature and technological background. This relation significantly aligns the verification process data, allowing for cheap and fast deals. More importantly, blockchain has tremendous potential to confirm the transition without requiring any central authorities. It is suggested by concerning trust, industries need to create a confidentiality platform for participants by improving features standards [80]. In addition, the rise of blockchain technology allows for the creation of digitized, shared, and trackable sets of data. 'Trustworthiness' in blockchain transparent features enable the information sharing to publicly accessible through its network infrastructure and ensure the credibility of data sets transaction [92].

(c) Traceability

Monitoring and verifying the production chain is critical to take certainty on ensuring how sources of contamination throughout the entire supply chain are processed. The innovation of blockchain has numerous points of interest to solving

traceability issues. Blockchain is programmable, dependable, and provides constant real-time operations throughout its traceability feature. The aim of traceability in the blockchain is for user to respond as fast as conceivable in any situation [93], as well as curbing unethical activity; improved services of sustainable practices; increased operational efficiency; increased coordination of the supply chain; and sensed market growth. The traceability feature makes blockchain a quality certification for industry, by local server and central management communication technologies. In addition, a study [94] reported that powerhouse retailers in Walmart cases, which have worked with IBM Food Trust to implement blockchain technology for their food chains for safety provenance. Such provenance is crucial to be emphasized to provide consumers, suppliers, retailers, manufacturers, and with more transparency on their goods and regulators to achieve the increase of ability to meet the pressure of the new custom for key participants. Blockchains help make the process more transparent by confirming each step with a block entry.

(d) Efficiency

Blockchain technology has attracted the attention of industry leaders with the potential to improve the efficiency and transparency of existing systems, particularly in system transactions and management data [95]. As a result, numerous industries include the government sector and early adopters made efforts to acquire blockchain capability and they explored blockchain as one of the important qualities necessary to generate innovative business models and evolve services significantly [96]. Countries around the world are seen committing to blockchain technology deployment, especially for the banking industry. For instance, HSBC implements the third wave of blockchain called Corda [97]. In terms of utilization, HSBC uses international trade that involves Malaysia and the United States. Corda offers security and efficiency in transactions. For example, Corda reduces the time for consensus financial transactions with bulk shipment to Malaysia of soya beans. This process normally takes about 5–10 days to organize and complete with numerous paper records. Since it is paperless, blockchain application resolves the issue in less than 24 h. Paper reconciliation is no longer relevant because every party is linked to the platform together and updates are available instantly.

(f) Reliability

Blockchain is more reliable traditional centralized due to improvement in the areas of incentive and sanction strategies, the process of manufacturing technology, restoration of the consensus network, and data restoration. As blockchain continues to be trending, many parties start to invest in the ecosystem, which makes the incentive for blockchain platforms massive by its nature. This popularity occurs due to blockchain decentralized nodes [98], in which there is no way for the government to find and tax the money stored in the blockchain, which means the sanction of business could be helpful. Proofs of retrievability technology of the blockchain store the core files in fragments, which to ensure reliability after the record has been

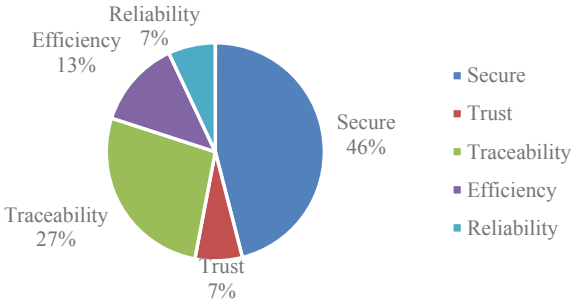


Fig. 2 Distribution of blockchain features adopted by industries

modified or added. As result, industries that rely on high reliability and security can attract more customers through the implementation of blockchain [10]. Indeed, blockchain technology is a distributed layer that prevents a single stage of malfunction situation.

In Fig. 2, it is shown that the highest percentage used in 12 industries in various contexts is "secure" with a total of 7 over 12 of 46%. AB InBev, Accenture, APL, and Kuehne + Nagel, Accredible, Spotify, North, and South Korea-international documents, Musicoin, CryptoKitties-visual arts, ANZ. and Westpac-commercial property are part of the industries that adopt blockchain due to the nature of transparency, payment bills, while transparency in document storage creating convenience for users to control their energy consumption. Besides, "traceability" is the second classification after "secure" with a total of 4 over 12 by having 27%. Industries such as Walmart, DNV GL, MyStory, AB InBev, Accenture, APL and Kuehne + Nagel and ANZ and Westpac use the supply chain platform to consolidate their statistics on the timeliness and quality of transactions provided to other participants of the chain. In addition, "efficiency" takes a total of 2 over 12 with 13%. HSBC in the banking sector and UN World Food Programme in Food and Agriculture Organization use blockchain to achieve efficient and accurate operations in a secured environment that enables them to deliver meaningful results in a safe and risk-free environment. Furthermore, "trust" and "reliability" classifications take a total of 1 over 12 with a percentage of 7%. Twiga Foods implement blockchain in terms of "trust" classification by helping reduce fraud as no party can add blockchain without the approval of the entire network that can be implemented in real-time. In a "reliability" classification, dredible-digital educational certificates ensure that the information is reliable and assures the representation of credentials with a record of qualitatively cryptographic credentials.

4.3 Results for RQ3

What are the recent trends of blockchain utilization in various industries? Based on the results for RQ1 and RQ2, valuable insights of use cases and classification have been identified. The insights help in establishing the annual trends of

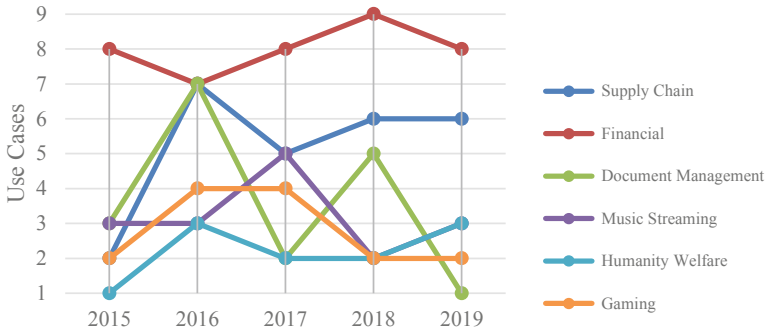


Fig. 3 Recent trends of blockchain use cases in industrial sectors

blockchain utilize in the industrial sector. Answering RQ3 supports other researchers and experts to concentrate further investigation in which such research is expected. Also, defining the current trends as shown in Fig. 3 allows us to understand the blockchain adoption evolution over the years.

We examined the research paper extracted from the data collection process to determine the emerging trends in blockchain advances. Four studies were found that support our analysis of blockchain use cases and feature classifications towards innovation [99–102]. In addition, all these research activities are becoming increasingly relevant in the current era of digitalization.

Figure 3 illustrates the trends of blockchain use cases, this use cases trend is in the form of graphs where it illustrates the percentage frequency adopted in the industrial sector. Within the year range, unstable agreement or encouragement can be observed for each use case in terms of percentage followed by the year. The most discussed and usage use cases in the review paper are financial with 80%, wherein a study [103] clarify that financial technology is offering a comprehensive range of services as such fees, exchange of local and foreign currency via P2P, investments, and fundraising. The lowest use cases in 2015 are humanity welfare, and this comes into an issue when a study [104] found that the cause is because of the coercion challenges that the problem regarding socioeconomic political systems is not being solved by the blockchain-based government. The changes in supply chain apparent spike can be seen rapidly increasing from 2015 to 2016, there's been a proof when a study [105] found that a developed series of logistics, methodologies, and principles to identify the supply chain with differ businesses strategy. Besides, a surge of increase in music streaming shows a change of percentage since for 2 years it just remains constant, where this investigation highlights [30] findings for the blockchain enhances the patterns to address and obtaining the music producer easier. In the year 2018, the use cases of gaming decreases, and the same goes to the remaining percentage in 2015 in constant horizontal from 2016 to 2017. In 2019 only human welfare use cases increase significantly, while documentation management is decreasing slightly as well financial, and the others remain constant as such gaming and supply chain. We can determine that the blockchain use cases

could be continued to expand by evolving the momentum platform of blockchain, new developments build an innovation industrial strategies and solution conventional blockchain-based networks.

Figure 4 illustrates the improvement that the blockchain brings that can be classified into 5 main categories: *security, trust, traceability, efficiency*, and *reliability*. The information in the graph is based on the 50 selected review papers, where a discussion of the author is reviewed to capture a frequent usage of blockchain classification. The highest classification discussed in 2015 is *trust* with 9 studies. One study [106] found that e-commerce websites are explicitly associated mostly with malicious third parties that control any purchases in the websites. Besides, the lowest classification in 2015 is traceability. A study [107] discovered that *traceability* of operations has become an important issue with the electronic currency world, this is why a domain controller with any payment method is formally instructed to increase exponentially in difficulty attack. In the year 2016, *trust,* and *security* increases usage in industry application. Blockchain technology is expected to revolutionize the industry via driving international trade transformation in the upcoming years [108]. There are decrease for most of the classification in 2017 excluding the classification of *efficiency* remain constant since 2016. Although, in 2018 all the 5 classifications increase slightly especially for the *reliability* feature. In the year 2018, and as reported by a study [109], reliable data becomes a tool required to maintain management plan as well as to establish a potential resource distribution management, industries thereby focus on building and upgrade the existing data system that ensures secure access controls. In 2019, only the *efficiency* decreases significantly, yet we can conclude that the trends show an improvement for the industrial sector because there is no significant fall in late 2018 and 2019. Blockchain research derives from multiple subject areas, which including science, businesses, and information system procedures for blockchain applications. Therefore, research and innovation started to focus on the solution for the biggest security threats in global, which include major potential threats that involve anonymous attackers who manipulated most of the network as such scam, denial-of-service attack, phishing. Initially, blockchain study started on industrial problems, thus more fields are expected to be associated as an effect of blockchain would reach improvement for the industrial, economic, and government.

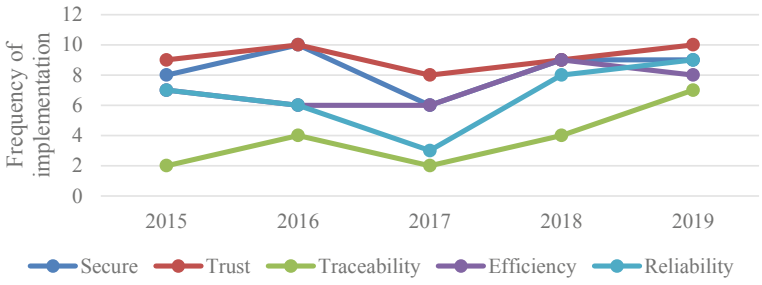


Fig. 4 Recent trends of implemented blockchain features in industrial sectors

5 Discussions

This section discusses the results and findings obtained from the review based on the formulated research questions. The discussions on each research question are presented as follows.

5.1 Overview of Blockchain Use Cases

In a total of 12 industries that this review observed, these industries adopted blockchain as the reliable mechanism for storing and processing their records securely. The aspect of trust seems to get attention significantly at the most outstanding factor throughout all use cases. A considerable part of the research of blockchain in industry emphasizes matters of privacy and security as well. In the use cases reviewed, blockchain is hugely useful for a transaction in various industries because of its competitiveness to numerous relations of stakeholders, customers, vendors, investors, and third parties. Such relations have significantly improved business processes, financial services, quality of service processes, expand the capability of overall use cases mentioned, and the reliability of data between businesses, services, and data exchange. Based on findings for RQ1, we learn that blockchain is not limited to improving processes in industries but also in the public sectors. Public sectors can leverage blockchain as a medium for social interaction exchange in improving the lives of citizens in the country.

5.2 Classifications of Blockchain Features

Based on the review results, five classifications of blockchain features are identified: security, trust, traceability, efficiency, and reliability. With this brief description, we also present issues and solutions to various industries as an understanding of blockchain's ability to overcome constraints. Therefore, we show that blockchain technology is not limited to any industry that wants to implement cryptocurrency into its systems. Instead, the ledger concept is distributed but not decentralized with trust and security features in the systems that can be customized with so many other techniques in the current business or industry sector. In addition, we have identified that 'security' is one of the key classifications for industry improvement. We also discovered that the industry adopts blockchain because the technology behind it promises secure transactions, where the industry processes a huge amount of data with the intention of implementing transparent storage through private transactions. This type of classification is expected to increase significantly, and it also remains to be a concern for various sectors in the long term because privacy and security are key aspects of establishing a decentralized transaction relationship with the industry.

5.3 Summarization of Blockchain Implementation Trends

Our findings on RQ3 indicates the blockchain capabilities in addressing industry issues with several interesting trends emerging. Our discussion shows that blockchain technology plays an important part in a growing industry by ensuring that its classification is maintained between use cases, and as well in future integration of newer implementation technologies such as cloud computing and IoT. The use of blockchain changes the performance of traditional data processes across various industries, also for the use of governments and other industries that have not yet used blockchain. At the same time, blockchain can be used by governments to improve efficiency in administration, as there are still public services that hesitate to adopt blockchain. With all the key classifications, the industry has recognized its strengths in using blockchain, this is due to many beneficial classifications adopted and more solutions to issues. Although blockchain has become more implemented, not all sectors use blockchain especially to the government when it does not improve the whole process because there are still shortcomings in terms of regulation. Nevertheless, there are several nations that are still amid integrating blockchain technology, while developed nations have consistently implemented blockchain into their operations and conducted high-level research.

6 Conclusions

Evidently, blockchain is a revolutionary technology that significantly transform the processes or approaches of many industries with innovative security features. In this paper, we provided the insights on significant blockchain advances and usage within information systems of many industries. Recent blockchain advances were analyzed to elicit notable applications, classifications, and emerging trends in blockchain technology. The presented classifications and characterizations of blockchain technology offer future researchers with fundamental comprehension to further support their investigation.

Our findings indicate that blockchain is becoming mainstream with various applications including securing supply chains, document management, entertainments, and public services. We learn that many notable enterprises have solidified the integration of blockchain unto their information systems or infrastructures. The concerns of privacy and security remain to be the utmost important aspects in integrating the blockchain in their systems. Based on the review, blockchain application features are classified into five categories based on their objectives: security, trust, traceability, efficiency, and reliability. Reinforcing security is the most sought-after feature in most of the blockchain applications and use cases presented in this review. Security is expected to remain the most important feature of blockchain integration, especially for future decentralized transactions. Our results presented broad applications of blockchain technology across many industry

domains, indicating growing interest and demand for blockchain integration. Despite the observed mainstream integration of blockchain, we found that the public sector has yet to fully leverage this technology. We infer that integrating blockchain into the public sector's existing systems or infrastructure also raises concern on policies and regulations.

Based on the findings, many new features are expected in future blockchain applications and innovations especially its adaptation in the public sectors. The existing works evidently emphasized the persistent transparency and conflict of interest issues in many public record transactions. Future studies can investigate the public's perception and acceptance of blockchain integration in public service information systems and infrastructures. The classifications of innovations brought by blockchain applications discussed in this paper are based on reported works in recent years. However, the advances of blockchain applications in the industry are realistically expeditious compared to reported works in literature. Therefore, future reviews can gather information from external sources like corporate reports, media, whitepapers, and patents to complement their findings. In addition, the presented discussions on use cases in this paper are based on notable blockchain applications recently reported in academic literature. Future studies can further enrich the blockchain use case discussions by gathering additional insights from industry experts via collaboration.

Acknowledgements This work is a part of a project submitted to Universiti Malaysia Sabah.

References

1. Lin Z, Li Y (2019) Design and implementation of classroom attendance system based on video face recognition. In: 2019 International conference on intelligent transportation, big data Smart City (ICITBS) (pp. 385–388). Presented at the 2019 international conference on intelligent transportation, big data Smart City (ICITBS). https://doi.org/10.1109/ICITBS.2019.00101
2. Senapati BR, Khilar PM, Sabat NK (2019) An automated toll gate system using VANET. In: 2019 IEEE 1st international conference on energy, systems, and information processing (ICESIP), pp 1–5. Presented at the 2019 IEEE 1st international conference on energy, systems and information processing (ICESIP). https://doi.org/10.1109/ICESIP46348.2019.8938380
3. Nanayakkara S, Perera S, Bandara D, Weerasuriya T, Ayoub J (2019) Blockchain technology and its potential for the construction industry. https://doi.org/10.6084/m9.figshare.10315331.v3
4. Yaga D, Mell P, Roby N, Scarfone K (2018) Blockchain technology overview. arXiv:1906.11078 [cs], NIST IR 8202. https://doi.org/10.6028/NIST.IR.8202
5. Zheng Z, Xie S, Dai H, Chen , Wang H (2017) An overview of blockchain technology: architecture, consensus, and future trends. In: 2017 IEEE international congress on big data (BigData Congress), pp 557–564. Presented at the 2017 IEEE international congress on big data (BigData Congress). https://doi.org/10.1109/BigDataCongress.2017.85
6. Chen J, Xue Y (2017) Bootstrapping a blockchain based ecosystem for big data exchange. In: 2017 IEEE international congress on big data (BigData Congress) (pp. 460–463).

Presented at the 2017 IEEE international congress on big data (BigData Congress), Honolulu, HI, USA: IEEE. https://doi.org/10.1109/BigDataCongress.2017.67
7. Bettín-Díaz R, Rojas AE, Mejía-Moncayo C (2018) Methodological approach to the definition of a blockchain system for the food industry supply chain traceability. In: Gervasi O, Murgante B, Misra S, Stankova E, Torre CM, Rocha AMAC, Ryu Y et al (eds) Computational science and its applications—ICCSA 2018, pp 19–33. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-95165-2_2
8. Pearce J (2019) Blockchain in academia: a literature review, 6
9. Gupta S, Schola M, Rao P, Pandey A (2019) Blockchain technology and use cases: open survey
10. Al-Jaroodi J, Mohamed N (2019) Blockchain in industries: a survey. IEEE Access 7:36500–36515. Presented at the IEEE Access. https://doi.org/10.1109/ACCESS.2019.2903554
11. Bodkhe U, Tanwar S, Parekh K, Khanpara P, Tyagi S, Kumar N, Alazab M (2020) Blockchain for industry 4.0: a comprehensive review. IEEE Access 8:79764–79800. Presented at the IEEE Access. https://doi.org/10.1109/ACCESS.2020.2988579
12. Hassan A, Mas'ud MZ, Shah WM, Abdul-Latip SF, Ahmad R, Ariffin A (2020) A systematic literature review on the security and privacy of the blockchain and cryptocurrency. OIC-CERT J Cyber Secur 2(1):1–17
13. Pournader M, Shi Y, Seuring S, Koh SCL (2020) Blockchain applications in supply chains, transport and logistics: a systematic review of the literature. Int J Prod Res 58(7):2063–2081. https://doi.org/10.1080/00207543.2019.1650976
14. Surjandy S, Widjaja M, Hidayanto A, Prabowo H (2019) The latest adoption blockchain technology in supply chain management: a systematic literature review. ICIC Exp Lett 13:913–920. https://doi.org/10.24507/icicel.13.10.913
15. Siddaway A, Wood A, Hedges L (2019) How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. Ann Rev Psychol 70. https://doi.org/10.1146/annurev-psych-010418-102803
16. Boell S, Cecez-Kecmanovic D (2015) On being 'systematic' in literature reviews in IS. J Inf Technol 30. https://doi.org/10.1057/jit.2014.26
17. Xiao Y, Watson M (2017) Guidance on conducting a systematic literature review. J Plan Edu Res 39, 0739456X1772397. https://doi.org/10.1177/0739456X17723971
18. Steiger E, de Albuquerque JP, Zipf A (2015) An advanced systematic literature review on spatiotemporal analyses of twitter data. Trans GIS 19(6):809–834. https://doi.org/10.1111/tgis.12132
19. Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering
20. Bürer MJ, de Lapparent M, Pallotta V, Capezzali M, Carpita M (2019) Use cases for blockchain in the energy industry opportunities of emerging business models and related risks. Comput Ind Eng 137:106002. https://doi.org/10.1016/j.cie.2019.106002
21. Li J, Greenwood D, Kassem M (2019) Blockchain in the built environment and construction industry: a systematic review, conceptual models and practical use cases. Autom Constr 102:288–307. https://doi.org/10.1016/j.autcon.2019.02.005
22. Bhalerao S, Agarwal S, Borkar S, Anekar S, Kulkarni N, Bhagwat S (2019) Supply chain management using blockchain. In: 2019 international conference on intelligent sustainable systems (ICISS), pp 456–459. Presented at the 2019 international conference on intelligent sustainable systems (ICISS). https://doi.org/10.1109/ISS1.2019.8908031
23. Gonczol P, Katsikouli P, Herskind L, Dragoni N (2020) Blockchain implementations and use cases for supply chains-a survey. IEEE Access 8:11856–11871. Presented at the IEEE Access. https://doi.org/10.1109/ACCESS.2020.2964880
24. Terzi S, Zacharaki A, Nizamis A, Votis K, Ioannidis D, Tzovaras D, Stamelos I (2019) Transforming the supply-chain management and industry logistics with blockchain smart contracts. In: Proceedings of the 23rd Pan-Hellenic conference on informatics, pp 9–14. Association for Computing Machinery, Nicosia, Cyprus. https://doi.org/10.1145/3368640.3368655

25. Bello G, Perez AJ (2019) Adapting financial technology standards to blockchain platforms. In: Proceedings of the 2019 ACM Southeast conference, pp 109–116.: Association for Computing Machinery, Kennesaw, GA, USA. https://doi.org/10.1145/3299815.3314434
26. Treleaven P, Brown R, Yang D (2017) Blockchain technology in finance. Computer 50:14–17. https://doi.org/10.1109/MC.2017.3571047
27. Youn S, Cho H (2019) Blockchain technology in finance industry. Trans Korean Inst Electr Eng 68:1601–1606. https://doi.org/10.5370/KIEE.2019.68.12.1601
28. Chavan S, Warke P, Ghuge S, Deolekar RV (2019) Music streaming application using blockchain. In: 2019 6th international conference on computing for sustainable global development (INDIACom), pp 1035–1040. Presented at the 2019 6th international conference on computing for sustainable global development (INDIACom)
29. De León I, Gupta R (2017) The impact of digital innovation and blockchain on the music industry. Inter-American Development Bank. https://doi.org/10.18235/0000881
30. O'Dair M, Beaven Z (2017) The networked record industry: How blockchain technology could transform the record industry. Strateg Chang 26(5):471–480. https://doi.org/10.1002/jsc.2147
31. Yuan C, Xu M, Si X, Li B (2017) Blockchain with accountable CP-ABE: how to effectively protect the electronic documents. In: 2017 IEEE 23rd international conference on parallel and distributed systems (ICPADS), pp 800–803. Presented at the 2017 IEEE 23rd international conference on parallel and distributed systems (ICPADS). https://doi.org/10.1109/ICPADS.2017.00111
32. Besançon L, Silva CFD, Ghodous P (2019) Towards blockchain interoperability: improving video games data exchange. In: 2019 IEEE international conference on blockchain and cryptocurrency (ICBC), pp 81–85. Presented at the 2019 IEEE international conference on blockchain and cryptocurrency (ICBC). https://doi.org/10.1109/BLOC.2019.8751347
33. Onik MH, Miraz MH, Kim C-S (2018) A recruitment and human resource management technique using blockchain technology for industry 4.0, 3 (6 pp.)-3 (6 pp.). https://doi.org/10.1049/cp.2018.1371
34. Ahram T, Sargolzaei A, Sargolzaei S, Daniels J, Amaba B (2017) Blockchain technology innovations. In: 2017 IEEE technology engineering management conference (TEMSCON), pp 137–141. Presented at the 2017 IEEE technology engineering management conference (TEMSCON). https://doi.org/10.1109/TEMSCON.2017.7998367
35. Niranjanamurthy M, Nithya BN, Jagannatha S (2019) Analysis of blockchain technology: pros, cons and SWOT. Clust Comput 22(6):14743–14757. https://doi.org/10.1007/s10586-018-2387-5
36. Zīle K, Strazdiņa R (2018) Blockchain use cases and their feasibility. Appl Comput Syst 23 (1):12–20. https://doi.org/10.2478/acss-2018-0002
37. Berg C, Davidson S, Potts J (2017) Blockchains Industrialise trust (SSRN Scholarly Paper No. ID 3074070). Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.3074070
38. Mohsin AH, Zaidan AA, Zaidan BB, Albahri OS, Albahri AS, Alsalem MA, Mohammed KI (2019) Blockchain authentication of network applications: taxonomy, classification, capabilities, open challenges, motivations, recommendations, and future directions. Comput Stand Interfaces 64:41–60. https://doi.org/10.1016/j.csi.2018.12.002
39. Narayanaswami C, Nooyi R, Govindaswamy SR, Viswanathan R (2019) Blockchain anchored supply chain automation. IBM J Res Dev 63(2/3):7:1–7:11. Presented at the IBM Journal of Research and Development. https://doi.org/10.1147/JRD.2019.2900655
40. Tan B, Yan J, Chen S, Liu X (2018) The impact of blockchain on food supply chain: the case of walmart. In: Qiu M (ed) Smart blockchain, pp 167–177. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-05764-0_18
41. Saberi S, Kouhizadeh M, Sarkis J, Shen L (2019) Blockchain technology and its relationships to sustainable supply chain management. Int J Prod Res 57(7):2117–2135. https://doi.org/10.1080/00207543.2018.1533261

42. Kshetri N (2018) 1 Blockchain's roles in meeting key supply chain management objectives. Int J Inf Manage 39:80–89. https://doi.org/10.1016/j.ijinfomgt.2017.12.005
43. Mills DC, Wang K, Malone B, Ravi A, Marquardt J, Badev AI et al (2016) Distributed ledger technology in payments, clearing, and settlement (SSRN Scholarly Paper No. ID 2881204). Social Science Research Network, Rochester, NY. https://doi.org/10.17016/FEDS.2016.095
44. Wu T, Liang X (2017) Exploration and practice of inter-bank application based on blockchain. In: 2017 12th international conference on computer science and education (ICCSE), pp 219–224. Presented at the 2017 12th international conference on computer science and education (ICCSE). https://doi.org/10.1109/ICCSE.2017.8085492
45. Wang R, Lin Z, Luo H (2019) Blockchain, bank credit and SME financing. Qual Quant 53 (3):1127–1140. https://doi.org/10.1007/s11135-018-0806-6
46. Hearn M, Brown RG (2019) Corda: a distributed ledger, 73
47. Graglia JM, Mellon C (2018) Blockchain and property in 2018: at the end of the beginning. Innov Technol Govern Globalization 12(1–2):90–116. https://doi.org/10.1162/inov_a_00270
48. Jirgensons M, Kapenieks J (2018) Blockchain and the future of digital learning credential assessment and management. J Teacher Educ Sustain 20(1):145–156. https://doi.org/10.2478/jtes-2018-0009
49. Yang J-H (2019) Applicability of blockchain based bill of lading under the Rotterdam rules and UNCITRAL model law on electronic transferable records (SSRN Scholarly Paper No. ID 3514414). Social Science Research Network, Rochester, NY. Retrieved from https://papers.ssrn.com/abstract=3514414
50. SA & SA (2017) Decentralized computing using blockchain technologies and smart contracts: emerging research and opportunities: emerging research and opportunities. IGI Global
51. García-Barriocanal E, Sánchez-Alonso S, Sicilia M-A (2017) Deploying metadata on blockchain technologies. In: Garoufallou E, Virkus S, Siatri R, Koutsomiha D (eds) Metadata and semantic research, pp 38–49. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-70863-8_4
52. Muzammal M, Qu Q, Nasrulin B (2019) Renovating blockchain with distributed databases: an open-source system. Futur Gener Comput Syst 90:105–117. https://doi.org/10.1016/j.future.2018.07.042
53. Loklindt C, Moeller M-P, Kinra A (2018) How blockchain could be implemented for exchanging documentation in the shipping industry. In: Freitag M, Kotzab H, Pannek J (eds) Dynamics in logistics, pp 194–198. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-74225-0_27
54. Halaburda H (2018) Blockchain revolution without the blockchain? Commun ACM 61 (7):27–29. https://doi.org/10.1145/3225619
55. Luu L, Chu D-H, Olickel H, Saxena P, Hobor A (2016) Making smart contracts smarter. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp 254–269. Association for Computing Machinery, Vienna, Austria. https://doi.org/10.1145/2976749.2978309
56. Ramachandran A, Kantarcioglu DM (2017) Using Blockchain and smart contracts for secure data provenance management. arXiv:1709.10000 [cs]. Retrieved from http://arxiv.org/abs/1709.10000
57. Udokwu C, Kormiltsyn A, Thangalimodzi K, Nor A (2018) An exploration of blockchain enabled smart-contracts application in the enterprise, 28
58. Arcos LC (2018) The blockchain technology on the music industry. Braz J Oper Prod Manag 15(3):439–443. https://doi.org/10.14488/BJOPM.2018.v15.n3.a11
59. Sitonio C, Nucciarelli A (2018) The impact of blockchain on the music industry. Calgary: International Telecommunications Society (ITS). Retrieved from https://www.econstor.eu/handle/10419/184968

60. Kshetr N, Voas J (2018) Blockchain in developing countries. IT Professional 20(2):11–14. Presented at the IT Professional. https://doi.org/10.1109/MITP.2018.021921645
61. Pol Català N (2018) Smart cities solutions for refugee camps: communication systems review to improve the conditions of refugees. Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-238672
62. Pisa M (2018) Reassessing expectations for blockchain and development. Innov Technol Govern Globalization 12(1–2):80–88. https://doi.org/10.1162/inov_a_00269
63. Min T, Wang H, Guo Y, Cai W (2019) Blockchain games: a survey. In: 2019 IEEE conference on games (CoG), pp 1–8. Presented at the 2019 IEEE Conference on Games (CoG). https://doi.org/10.1109/CIG.2019.8848111
64. Cascio WF, Montealegre R (2016) How technology is changing work and organizations. Annu Rev Organ Psych Organ Behav 3(1):349–375. https://doi.org/10.1146/annurev-orgpsych-041015-062352
65. Paik H-Y, Xu X, Bandara HMND, Lee SU, Lo SK (2019) Analysis of data management in blockchain-based systems: from architecture to governance. IEEE Access 7:186091–186107. Presented at the IEEE Access. https://doi.org/10.1109/ACCESS.2019.2961404
66. Komiya K, Nakajima T (2019) Increasing motivation for playing blockchain games using proof-of-achievement algorithm. In: Fang X (ed) HCI in games, pp 125–140. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-22602-2_11
67. Serada A, Sihvonen T. Harviainen JT (2020) CryptoKitties and the new ludic economy: how blockchain introduces value, ownership, and scarcity in digital gaming. Games Culture 1555412019898305. https://doi.org/10.1177/1555412019898305
68. Joshi AP, Han M, Wang Y (2018) A survey on security and privacy issues of blockchain technology. Math Found Comput 1(2):121. https://doi.org/10.3934/mfc.2018007
69. Li X, Jiang P, Chen T, Luo X, Wen Q (2020) A survey on the security of blockchain systems. Futur Gener Comput Syst 107:841–853. https://doi.org/10.1016/j.future.2017.08.020
70. Şahan S, Ekici AF, Bahtiyar Ş (2019) A multi-factor authentication framework for secure access to blockchain. In: Proceedings of the 2019 5th international conference on computer and technology applications, pp 160–164. Association for Computing Machinery, Istanbul, Turkey. https://doi.org/10.1145/3323933.3324083
71. Taylor PJ, Dargahi T, Dehghantanha A, Parizi RM, Choo K-KR (2019) A systematic literature review of blockchain cyber security. Digital Commun Netw. https://doi.org/10.1016/j.dcan.2019.01.005
72. Zhang R, Xue R, Liu L (2019) Security and privacy on blockchain. ACM Comput Surv 52(3):51:1–51:34. https://doi.org/10.1145/3316481
73. Mann S, Potdar V, Gajavilli RS, Chandan A (2018) Blockchain technology for supply chain traceability, transparency and data provenance. In: Proceedings of the 2018 international conference on blockchain technology and application, pp 22–26. Xi'an, Association for Computing Machinery, China. https://doi.org/10.1145/3301403.3301408
74. Song JM, Sung J, Park T (2019) Applications of blockchain to improve supply chain traceability. Procedia Comput Sci 162:119–122. https://doi.org/10.1016/j.procs.2019.11.266
75. Xu X, Rahman F, Shakya B, Vassilev A, Forte D, Tehranipoor M (2019) Electronics supply chain integrity enabled by blockchain. ACM Trans Design Autom Electron Syst 24(3):31:1–31:25. https://doi.org/10.1145/3315571
76. Galvez JF, Mejuto JC, Simal-Gandara J (2018) Future challenges on the use of blockchain for food traceability analysis. TrAC Trends Anal Chem 107:222–232. https://doi.org/10.1016/j.trac.2018.08.011
77. Jeppsson A, Olsson O (2017) Blockchains as a solution for traceability and transparency. Retrieved from http://lup.lub.lu.se/student-papers/record/8919957
78. Abebe E, Behl D, Govindarajan C, Hu Y, Karunamoorthy D, Novotny P et al (2019) Enabling enterprise blockchain interoperability with trusted data transfer (Industry Track). In: Proceedings of the 20th international middleware conference industrial track, pp 29–35. Association for Computing Machinery, Davis, CA, USA. https://doi.org/10.1145/3366626.3368129

79. Fremont V, Jonathan GM (2018) Can blockchain technology solve trust issues in industrial networks?
80. Shin DDH (2019) Blockchain: the emerging technology of digital trust. Telematics Inform 45:101278. https://doi.org/10.1016/j.tele.2019.101278
81. Liu B, Mei S (2019) Research on investment efficiency of listed companies based on DEA blockchain concept. In: Proceedings of the 2019 the world symposium on software engineering, pp 93–97. Association for Computing Machinery, Wuhan, China. https://doi.org/10.1145/3362125.3362136
82. Maydanova S, Ilin I (2018) Blockchain as a tool for shipping industry efficiency increase
83. Muzylyov D, Shramenko N (2020) Blockchain technology in transportation as a part of the efficiency in industry 4.0 strategy, pp 216–225. https://doi.org/10.1007/978-3-030-40724-7_22
84. Peng Z, Wu H, Xiao B, Guo S (2019) VQL: providing query efficiency and data authenticity in blockchain systems. In: 2019 IEEE 35th international conference on data engineering workshops (ICDEW), pp 1–6. Presented at the 2019 IEEE 35th international conference on data engineering workshops (ICDEW). https://doi.org/10.1109/ICDEW.2019.00-44
85. Baralla G, Ibba S, Marchesi M, Tonelli R, Missineo S (2019) A blockchain based system to ensure transparency and reliability in food supply chain. In: Mencagli G, Heras DB, Cardellini V, Casalicchio E, Jeannot E, Wolf F et al (eds) Euro-Par 2018: parallel processing workshops, pp 379–391. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-030-10549-5_30
86. Lo SK, Xu X, Staples M, Yao L (2020) Reliability analysis for blockchain oracles. Comput Electr Eng 83:106582. https://doi.org/10.1016/j.compeleceng.2020.106582
87. Kher R, Terjesen S, Liu C (2020) Blockchain, Bitcoin, and ICOs: a review and research agenda. Small Bus Econ. https://doi.org/10.1007/s11187-019-00286-y
88. Casino F, Dasaklis TK, Patsakis C (2019) A systematic literature review of blockchain-based applications: current status, classification and open issues. Telematics Inform 36:55–81. https://doi.org/10.1016/j.tele.2018.11.006
89. Puthal D, Malik N, Mohanty SP, Kougianos E, Yang C (2018) The blockchain as a decentralized security framework [Future Directions]. IEEE Consumer Electron Mag 7 (2):18–21. Presented at the IEEE consumer electronics magazine. https://doi.org/10.1109/MCE.2017.2776459
90. Krause E, Velamuri V, Burghardt T, Nack D, Schmidt M, Treder T-M (2016) Blockchain technology and the financial services market state-of-the-art analysis
91. Mou J, Shin D-H, Cohen JF (2017) Trust and risk in consumer acceptance of e-services. Electron Commer Res 17(2):255–288. https://doi.org/10.1007/s10660-015-9205-4
92. Seebacher S, Schüritz R (2017) Blockchain technology as an enabler of service systems: a structured literature review. In: Za S, Drăgoicea M, Cavallari M (eds) Exploring services science, pp 12–23. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-56925-3_2
93. Hastig G, Sodhi M (2019) Blockchain for supply chain traceability: business requirements and critical success factors. Prod Oper Manag. https://doi.org/10.1111/poms.13147
94. Kamath R (2018) Food traceability on blockchain: Walmart's Pork and Mango Pilots with IBM. J Brit Blockchain Assoc 1:1–12. https://doi.org/10.31585/jbba-1-1-(10)2018
95. Yuan Y (2018) Blockchain and cryptocurrencies: model, techniques, and applications. IEEE Trans Syst Man Cybern Syst 48:1421–1428. https://doi.org/10.1109/TSMC.2018.2854904
96. Alketbi A, Nasir Q, Abu Talib M (2020) Novel blockchain reference model for government services: Dubai government case study. Int J Syst Assur Eng Manag. https://doi.org/10.1007/s13198-020-00971-2
97. Mohanty D (2019) The Latest from Corda: cordite, Corda enterprise, and settler. In: Mohanty D (ed) R3 Corda for architects and developers: with case studies in finance, insurance, healthcare, travel, telecom, and agriculture, pp 199–205. Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-4529-3_12

98. Ølnes S, Ubacht J, Janssen M (2017) Blockchain in government: benefits and implications of distributed ledger technology for information sharing. Gov Inf Q 34(3):355–364. https://doi.org/10.1016/j.giq.2017.09.007
99. Beck R, Müller-Bloch C (2017) Blockchain as radical innovation: a framework for engaging with distributed ledgers as incumbent organization. In: Hawaii international conference on system sciences 2017 (HICSS-50). Retrieved from https://aisel.aisnet.org/hicss-50/os/practice-based_research/3
100. Duy PT, Hien DTT, Hien DH, Pham V-H (2018) A survey on opportunities and challenges of Blockchain technology adoption for revolutionary innovation. In: Proceedings of the ninth international symposium on information and communication technology, pp 200–207. Association for Computing Machinery, Danang City, Viet Nam. https://doi.org/10.1145/3287921.3287978
101. Kim S, Park H, Lee J (2020) Word2vec-based latent semantic analysis (W2V-LSA) for topic modeling: A study on blockchain technology trend analysis. Exp Syst Appl 152:113401. https://doi.org/10.1016/j.eswa.2020.113401
102. Wang S, Ouyang L, Yuan Y, Ni X, Han X, Wang F-Y (2019) Blockchain-enabled smart contracts: architecture, applications, and future trends. IEEE Trans Syst Man Cybern Syst 49 (11):2266–2277. Presented at the IEEE Transactions on Systems, Man, and Cybernetics: Systems. https://doi.org/10.1109/TSMC.2019.2895123
103. Peters G, Panayi E, Chapelle A (2015) Trends in cryptocurrencies and blockchain technologies: a monetary theory and regulation perspective (SSRN Scholarly Paper No. ID 3084011). Social Science Research Network, Rochester, NY. Retrieved from https://papers.ssrn.com/abstract=3084011
104. Atzori M (2015) Blockchain technology and decentralized governance: is the state still necessary? (SSRN Scholarly Paper No. ID 2709713). Social Science Research Network, Rochester, NY. https://doi.org/10.2139/ssrn.2709713
105. Vukatana K, Sevrani K, Hoxha E (2016) Wine traceability: a data model and prototype in Albanian context. Foods 5(1):11. https://doi.org/10.3390/foods5010011
106. Crosb M (2016) BlockChain technology: beyond bitcoin (2), 16
107. Tasca P (2015) Digital currencies: principles, trends, opportunities, and risks. SSRN Electron J. https://doi.org/10.2139/ssrn.2657598
108. Underwood S (2016) Blockchain beyond bitcoin. Commun ACM 59(11):15–17. https://doi.org/10.1145/2994581
109. Wierzbowski P (2018) Application of blockchain technology in information management in supply chains. Transp Econ Logistics 78:179–191. https://doi.org/10.26881/etil.2018.78.15

Improved Group-Based Signaling Scheme for Mobile Nodes in the Internet of Things Networks

Safwan M. Ghaleb, Mustafa Man, Baraq Ghaleb, Faizah Aplop, Mukhtar Ghaleb, and Ismail Ahmed Al-Qasem Al-Hadi

Abstract One of the most important challenge in the mobility management protocols is maintaining the MN's IP connectivity and its accessibility during its movement. Although mobility protocols have come a long way in solving this issue in IP-WSNs, they still suffer from system performance degradation during the handoff process (e.g., long handover time and heavy signaling), which is not allowable specially in the real-time applications. In this article, Enhanced Mobility Cluster Sensor Proxy Mobile IPv6 (EM-CSPMIPv6) scheme is proposed to solve these issues via keeping the MN reachable while minimizing the time of inaccessible network. EM-CSPMIPv6 introduces a fast handover scheme through applying two mechanisms. In the NON-Remote clustering (CN-MN) mechanism, this done through clustering the MNs that are moving as a group to register them in advance before their actual handoff. In the Remote-based Clustering (CR-MN) mechanism, we suggest to simultaneously manipulate mobility-related signaling of the MNs arriving to the target Access Rout (AR) at the same time to reduce the bandwidth waste and the signaling cost. To verify our proposed scheme, an extensive simulation is performed through comparing EM-CSPMIPv6 with FMIPv6 protocol and GB-FH method. The results show that the EM-CSPMIPv6 increases the performance of system in terms of handoff latency, signaling cost and bandwidth usage when compared to the counterpart.

Keywords IP-WSN · Mobility managements · Fast mobile IPv6 · Mobile IPv6 · Handover latency

S. M. Ghaleb (✉) · M. Man · F. Aplop · I. A. A.-Q. Al-Hadi
Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu (UMT), 21030 Kuala Nerus, Terengganu, Malaysia

B. Ghaleb
School of Computing At Edinburgh, Napier University, 10 Colinton Road, Edinburgh E10 5DT, UK

M. Ghaleb
School College of Sciences and Arts, University of Bisha, Alnamas 61977, Saudi Arabia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_34

1 Introduction

Designing a protocol for connecting the Wireless Sensor Network (WSN) with the IP networks is become a necessity due to the emergence of the Internet of Thing (IoT) and the machine self-communication without any user involvement. WSNs are used to gather the information and transmit it to a central processing unit for further processing [1–3]. To make the WSNs accessible through the Internet, IPv6 and 6LoWPAN have been introduced into WSNs for the purpose of locating and connecting them to the network. The integration of IPv6 and 6LoWPAN with WSNs has given the birth for a multitude of applications such as military operations, healthcare, smart home, surveillance applications, monitoring applications and Intelligent Transport System (ITS).

The exponential growth of mobile devices in WSNs applications has emphasized the need for developing an efficient and seamless handover protocols. Hence, several protocols have been proposed aiming to facilitate the mobility in WSNs. These mobility protocols have been categorized into host-based and network-based protocols. Protocols such as Mobile IP Version 6 (MIPv6) [4], Fast Mobile Internet Protocol Version 6 (FMIPv6) [5] and Hierarchical Mobile IPv6 (HMIPv6) [6] belong to the host-based which our concern in this paper. The host-based protocols are totally depending on involvement of the Mobile Node (MN) in the handover process [7]. Thus, The MN complexity in the host-based protocol is high due to its involvement in the control signaling messages to configure its Care of Address (CoA) through sending Binding Update (BU) and receiving Binding Acknowledgment (BA) messages. This may result in system performance degradation, which is not acceptable in the real-time sessions.

In the urban cities, the MNs frequently are triggered their handover in a short time. This is happening once the vehicles either private or public contain several MNs and travel with a high velocity. Accordingly, if the MNs perform their handover individually this will cause a heavy signaling and high bandwidth cost, which in turn leads to high handover latency. If the MNs being grouped the signaling will be reduced. Thus, In this work, an Enhanced Mobility CSPMIPv6 (EM-CSPMIPv6) scheme is proposed to deal with multiple nodes as one node in order to increase the Quality of Service (QoS) of the mobility protocols through grouping several MNs into a cluster with cluster head (MNH). The proposed EM-CSPMIPv6 is aimed to reduce the heavy signaling required in the host-based which in turn leads to reduce the MN's complexity. To show the superiority of Our EM-CSPMIPv6 scheme, a comparison with the host-based protocols that are represented by FMIPv6 protocol and GB-FH method through performing simulation experiments is made in this paper.

Despite of the efficiency of the host-based protocols, they still suffer from some issues including signaling cost, handover latency, bandwidth usage, and un-optimized path. In addition, most of the prior mobility management protocols deal with a single MN during the handover process. Consequently, a method, named Group-based Fast Handover (GB-FH), has been proposed in [8] in order to

predict a multiple node movement to lower the signaling cost and minimize the handover delay time for multiple nodes. However, GB-FH have several limitations that negatively affect the performance system, which makes it unsuitable for the critical applications that are mentioned in the next section.

This article presents an efficient scheme to solve issues of high signaling cost, handover delay time and bandwidth waste associated with prior protocols.

In this article, several dominant contributions are introduced as follows:

- An extensive preview of the mobility management protocols, which are used to facilitate the MNs movement without service disruption.
- Identification of demerits of FMIPv6 and GB-FH, which represents the underlying motivation of our proposed EM-CSPMIPv6 scheme.
- A new scheme to improve the effectiveness of the mobility protocols.
- Extensive simulations to evaluate our proposed scheme in comparison with host-based protocols.

The rest of this paper is organized as follows: Section 2 discusses the related works with highlighting their cons and pros. Section 3 proposes the system model which is employed as the environment for our EM-CPMIPv6 scheme. In Sect. 4 demonstrates the EM-CSPMIPv6 methodology within the host-based protocols environments. In Sect. 5, the performance evaluation of EM-CPMIPv6 Scheme is discussed. Finally, Sect. 6 summarized the study contributions.

2 Mobility-Related Study

Mobility is an effective technique that leads to discovering IoT abilities to achieve "anywhere, anytime" communications. Mobility management in IoT networks involves two important functions: location management and handoff management [9]. Location management is a function used to track the MN's location. Handoff management is a function used to maintain MN's connection during its roaming [9, 10].

Even though WSNs are stationary networks mainly, there are a considerable number of scenarios where the mobility is the norm. Mobility management protocols play an important role in the current WSN applications such as in military applications, healthcare and mobile vehicles. This is because the users need to retrieve the gathered information from the sensors, while they are in motion. Host-based is originally designed to minimize the inaccessible network time in the consecutive connections as well as to guarantee MN reachability.

MIPv6 has been proposed to support the mobility management during the MN movement. Home Agent (HA) in the MIPv6 is responsible to redirect the packets to the new MN's Care of Address (CoA) rather than to the old MN' CoA, when HA receives the new CoA from the MN. Thus, the packets are continuously delivered to the roaming MN without interruption. Although, MIPv6 supports mobility, it needs

a long time to finish the handover operations, which increases the handoff delay time [11–14].

Many recent mechanisms have been proposed with the main aim to enhance the efficiency in terms of handoff delay and according to the application requirements. The researchers in [14–16] focus on the data link layer to decrease the handover latency and the researchers in [17, 18] focus on reducing the signaling cost. Network Mobility (NEMO) protocols have been proposed in [19–24] to support the mobility for a group of MNs. In the NEMO protocols, the term Mobile Router (MR) is used to refer to the entity that handles the mobility-related signaling on the behalf of its associated with MNs. Thus, the handoff delay time is shortened and the overhead of the signaling is reduced. NEMO protocols are useful in many applications such as: military applications, vehicle movement and patient healthcare.

Another work in the host-based protocol is proposed by Internet Engineering Task Force (IETF), named as Fast Mobile Internet Protocol version 6 (FMIPv6). This work is aimed to minimize the handover delay time. The MN in this protocol is involved during the handover process to accelerate the handover operations through triggering the L3 handover before completing the L2 handover. In other words, the MN requests a new CoA through exchanging Router Solicitation (RS) and Router Advertisement (RA) with Previous Access Router (PAR). This happens when the MN movement is detected toward the New Access Router (NAR), as shown in Fig. 1. The MN sends Fast Binding Update Message (FBU) to inform the PAR that the packets must be forwarded to the NAR. Then PAR and NAR form a bi-directional tunnel between them through exchanging Handover Initiate (HI) and Handover Acknowledgment (HAck) messages. A tunnel between PAR and NAR is setup to forward the packets between them. Fast Binding Acknowledgment (FBAck) message is sent by the PAR to both MN and NAR access links to confirm the handover preparation.

Despite of the benefits offered by previous host-based protocols (e.g., reducing the handover latency and signaling cost), the required level of QoS still is not achieved. This is because the protocols consider only a single MN in their domain, which in turn increases the signaling cost and wastes the bandwidth. Unlike the network-based protocols which manipulate the handover with various ways such as grouping signaling, Storing mechanism [25] and load balancing technique [26]. Subsequently, the work in [8] develops approach, named Group-based Fast Handover (GB-FH), to enhance system performance of FMIPv6. This is achieved by selecting one MN, to act as a guide for several neighboring nodes. The main aim of this work is to predict the MNs movement and perform a pre-registration process for them. The MN guide prepares itself for the handover before the actual handover processes are performed. The GB-FH method reduces the time needed for handover initiation, messages signaling, and handover latency compared to FMIPv6 according to their analytical analysis and simulation results. The GB-FH method signaling, including the MN registration operations are illustrated as follows:

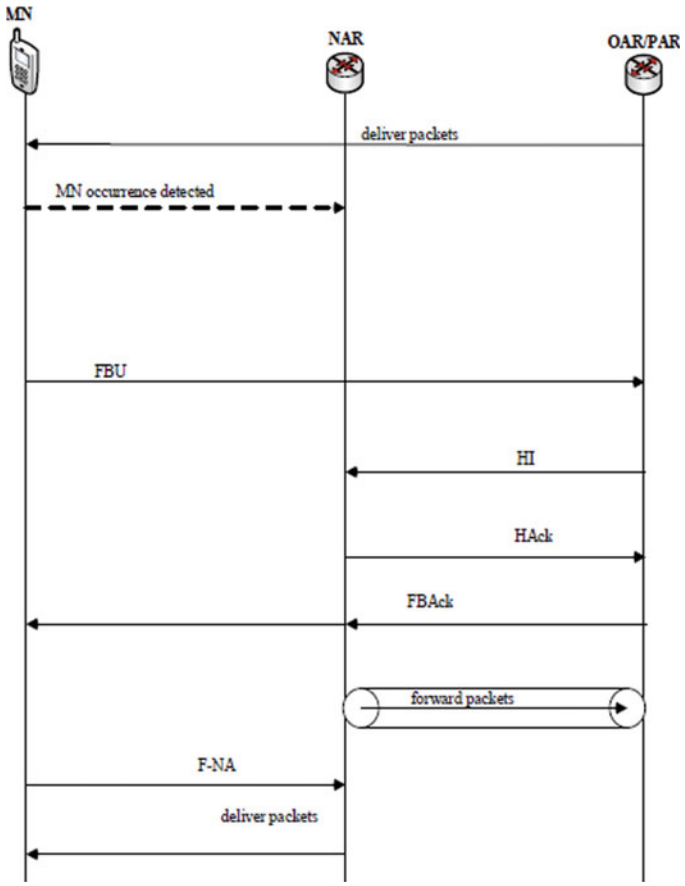


Fig. 1 Message flow diagram of the FMIPv6 protocol (Predictive scenario)

1. The MN guide sends a B-Reg-Join broadcast message to its neighboring MNs asking them to be joint. This is performed when the MN guide is still connected to its Old AR (OAR) properly and its movement is not detected yet by the NAR.
2. When the MN guide detects NAR signaling, it rejects any accept join messages arriving afterward.
3. Once the MN guide receives its neighbors' Acc-Join messages, which include their MAC addresses, it places them in one list as a group.
4. In the actual handover process, if the guide node has detected a signal, whose power is more than a pre-defined threshold, it sends the FBU message in a broadcast manner to its OAR including the MAC addresses of the neighboring MNs who joined successfully. The MN guide sends B-FBU in broadcast to trigger its actual handover and inform the neighboring MNs about their successful joint.

5. Upon receiving the FBU message by the OAR, the OAR specifies the target network and then sends a HI message to the NAR including the CoAs of the MNs and the MN guide identification. The HI message not only establishes a bi-directional tunnel, but also asks the NAR to check the CoAs.
6. Upon receiving the HI message, the NAR checks the CoAs and responses to the OAR with HAck message.
7. Once the HAck message is received by OAR successfully, it sends FBAck message in a broadcast manner to the MN guide, NAR and the neighboring MNs who joined successfully.
8. When the MN guide receives the FBAck message successfully, it starts data link layer handover in order to connect with the NAR.
9. The neighboring MNs inform their successful join upon receiving the FBAck message, which contains their new CoAs.

Despite all the benefits that have been introduced by the GB-FH method, GB-FH method still suffers from several limitations, as illustrated as follows.

- MN guide communicates with its neighboring MNs that are coming only from same network, which limits the number of predictable MNs in the system.
- MN guide rejects the request messages from the neighboring MNs, when its actual handover begins with the new network. This makes the system improvement useless, when the MNs come in a sequential order in the overlapping area (sparse area).
- MN guide does not consider the de-registration messages, which may happen before registering the MN in the new network that may increase the handover latency time.
- MN guide accepts the request message from the neighboring MNs, which move in another direction such as vertical direction that may result in increasing false predictable MNs.
- MN guide has a high probability to change its direction after requesting the neighboring MNs to join it. This is due to the performance of an MNs grouping outside the overlapping area by the MN guide, which negatively affects the system.
- MN guide uses the same packet for informing the new network for registering itself and its neighboring nodes. This may negatively impact the registration timing, which is not acceptable in the critical session connection.

Another interested work is proposed by [27] to reduce the handover latency through applying Bursty Multi-node Handover scheme for mobile internet using the partially Distributed Mobility Management (BMH–DMM) architecture. The Media Independent Handovers (MIH) framework is integrated in their work to achieve seamless handover for several MNs instead of one MN. According to their performance analysis results, the network load, signaling cost, packet loss and handoff delay are reduced comparing with that of using the individual handover method.

3 System Models

This section explains the system models used for our proposed EM-CSPMIPv6 scheme. These models are explained as follows.

- An infrastructure-based vehicular network within host-based domain is used, as shown in Fig. 2. Several Access Routers (ARs) are used to cover the area along the roads to preserve the roaming MNs connections.

In this model, the vehicles are moving in the opposite direction and they have one or more MNs that are communicating with other nodes outside/inside their network.

The coverage overlapping is demonstrated in our system model within two ARs/MAGs (MAGs mean ARs) coverage area, as shown in Fig. 3. This overlapping is divided into two coverage areas. This is done through applying threshold criteria. The notations and specification are presented as follows.

- R refers to the radius of circular which is a covered area by the AR.
- S_{th} or S_{ath} refers to the minimum threshold value that the MNs can receive the Received Signal Strength (RSS) to successfully joining with AR_2.
- S_{min} refers to the minimum RSS threshold value, in which the MNs can form a group from the neighboring MNs to do pre-registration signaling for all the MNs who successfully joint the grouped.

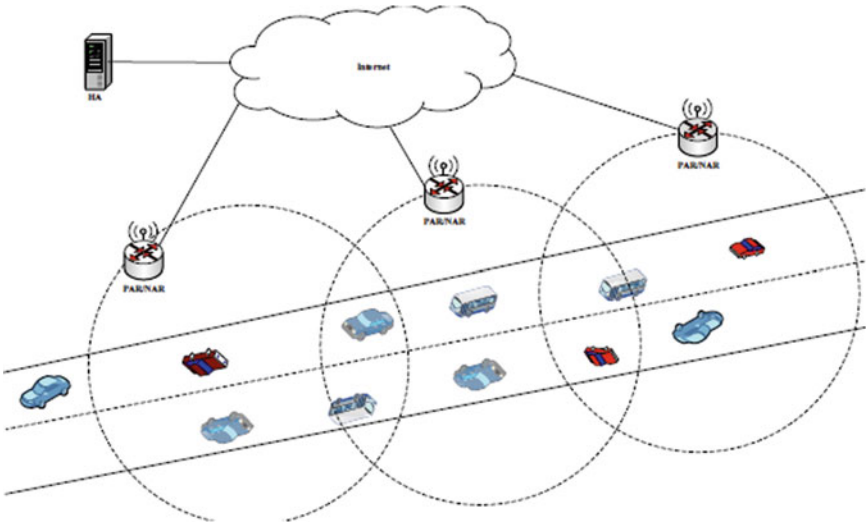


Fig. 2 System Architecture of EM-CSPMIPv6 scheme in Host-based domain

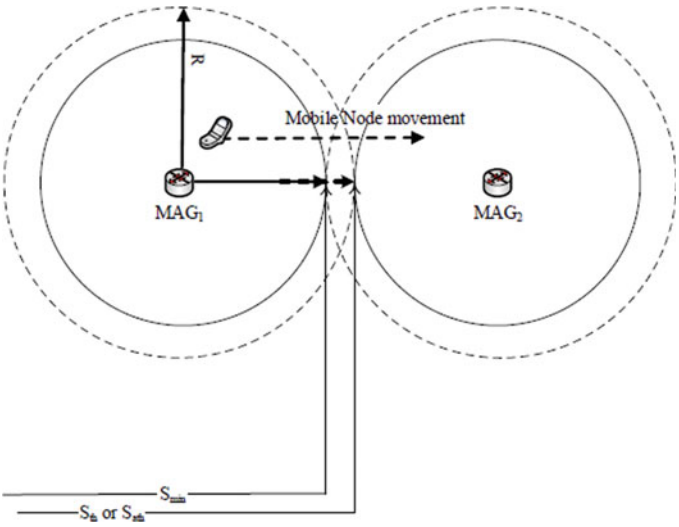


Fig. 3 Overlapping area within the EM-CSPMIPv6 system model

4 The Proposed EM-CSPMIPv6 Scheme

In this section, we will discuss and present the proposed EM-CSPMIPv6 scheme, which enhances the mobility via minimizing handover time and signaling cost within the mobility management protocols. The proposed scheme is implemented in the host-based domains. The host-based protocols are represented by the FMIPv6 protocol and GB-FH method explained in previous section. Two mechanisms are suggested and proposed to achieve the improvements. In CN-MN mechanism, the MNs that move closely from each other between two networks within the FMIPv6 domain are grouped. This mechanism is implemented to overcome the issues related the GB-FH method that employed the MNs grouping using the FMIPv6 principle. In the CR-MN mechanism, the MNs that are simultaneously triggered their mobility to the same AR are processed together to reduce the signaling cost as well as the bandwidth usage. EM-CSPMIPv6 considers the FMIPv6's entities (e.g., AR) without any modification in both mechanisms. In this section, the proposed EM-CSPMIPv6 scheme is discussed within the host-based environment.

4.1 EM-CSPMIPv6 Scheme from Host-Based Perspective

Figure 4 shows the message flow diagram of our EM-CSPMIPv6 scheme within FMIPv6 domain based on the CN-MN mechanism. As shown in Fig. 4, the MN measures its RSS continuously to take a decision for performing clustering steps in

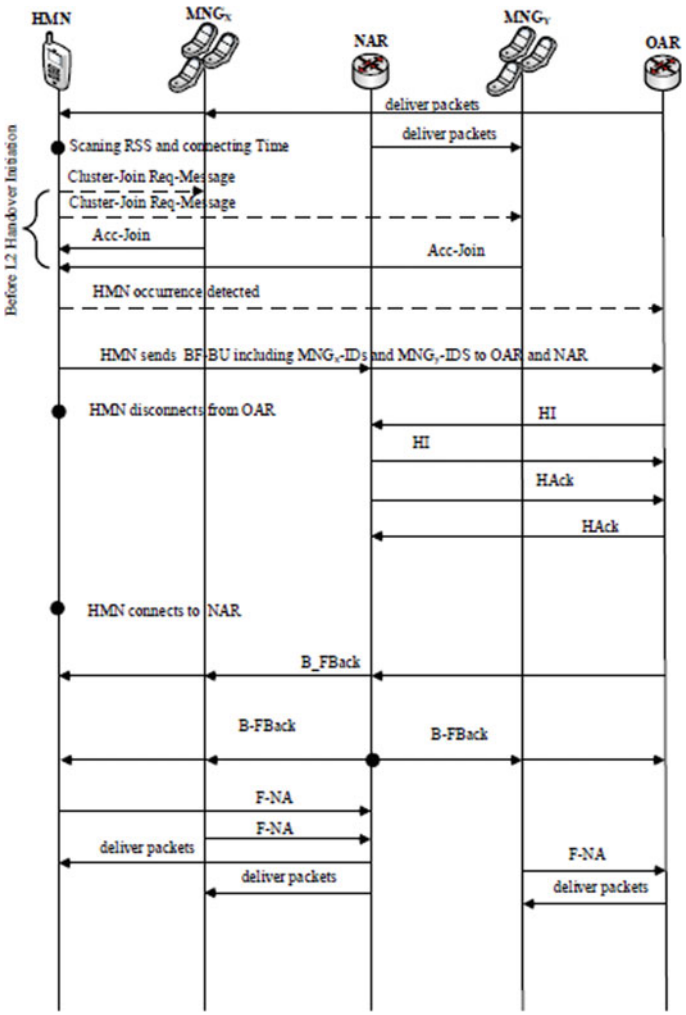


Fig. 4 The message flow diagram for EM-CSPMIPv6 scheme within FMIPv6 domain from CN-MN mechanism

order to group the neighboring MNs before its actual handover is initiated. The steps of clustering the neighboring MNs are performed in the same way that are performed by the EM-CSPMIPv6 scheme in CSPMIPv6 domain, which are discussed in the above section. After completing the clustering process, the HMN performs some necessary steps to pre-register their member MNs, as follows.

- When the HMN is detected by the NAR and the $ S_{th} $ threshold is reached, a FBU is sent by the HMN to both NAR and OAR including addresses of the successfully joined MNs.

- Then HI and HAck messages are exchanged between NAR and OAR to tunnel the packets from the old MN'sCoA to the new MN'sCoA as well as to check the CoAs uniqueness.
- Once the OAR and NAR receives the HAck message from NAR and OAR respectively, F-Back is sent by old HMN'sCoA to the new HMN'sCoA and neighboring MNs, as depicted in Fig. 4.
- When the HMN receives F-Back message simultaneously, the HMN starts a link layer disconnection from the OAR and binding to the NAR through sending Fast Neighbor Advertisement (FNA) in order to inform its presence to the NAR. After that, the NAR forwards the packets to the presence MN.
- When the neighboring MN, which is successfully joined, is detected by the NAR, it sends F–NA message to inform its presence to the NAR and starts a new L2 connection with the new CoA.

The CR-MN mechanism occurs when several MNs are attached to NAR at the same time and their mobility-related signaling are processed simultaneously by the NAR to mitigate the bandwidth burden, as shown in Fig. 5. When the MNs movement are detected by the NAR and the RSS S_{th} threshold is reached, the NAR adds the IDs of the expected MNs to a list created for grouping the MNs that are expected to move from their OAR to the NAR around the same time. Consequently, the mobility-related signaling is performed simultaneously for several MNs, which in turn reduces greatly the HI and HAck messages between the OAR and NAR. This comes because of including several MNs addresses in the HI and HAck messages in order to register them and check their CoAs at the same time.

Note that our proposed scheme has several benefits regarding the E-CSPMPv6 scheme as illustrated as follows.

- In Fig. 6 the diamond interchange road is considered if it exists in the overlapping area between ARs. This is done via applying a time threshold value. Every MN, which receives a Req-join message will not reply to this message if the time period connected to its serving AR is less than the time threshold value. This definitely leads to reduce the false prediction by preventing the newer joining MNs to their serving ARs from joining the prediction MNs list.
- De-registration processes are eliminated from the handover time via sending the lists created by the HMN to both ARs at a proper time (serving AR, new AR). This leads to increase the system accuracy as well as the number of the MNs in the list prediction, which in turn reduces the handover latency, signaling cost and bandwidth waste.
- In order to joining MNs as much as possible, the HMN maintain the ability to receive the requests messages from the MNs after even completing its own registration process till reached the pre-defined threshold.
- The HMN has two packets during its registration, one for its normal registration and the other one for register the neighboring MNs. This leads to eliminate the time effect via registering several MNs simultaneously.

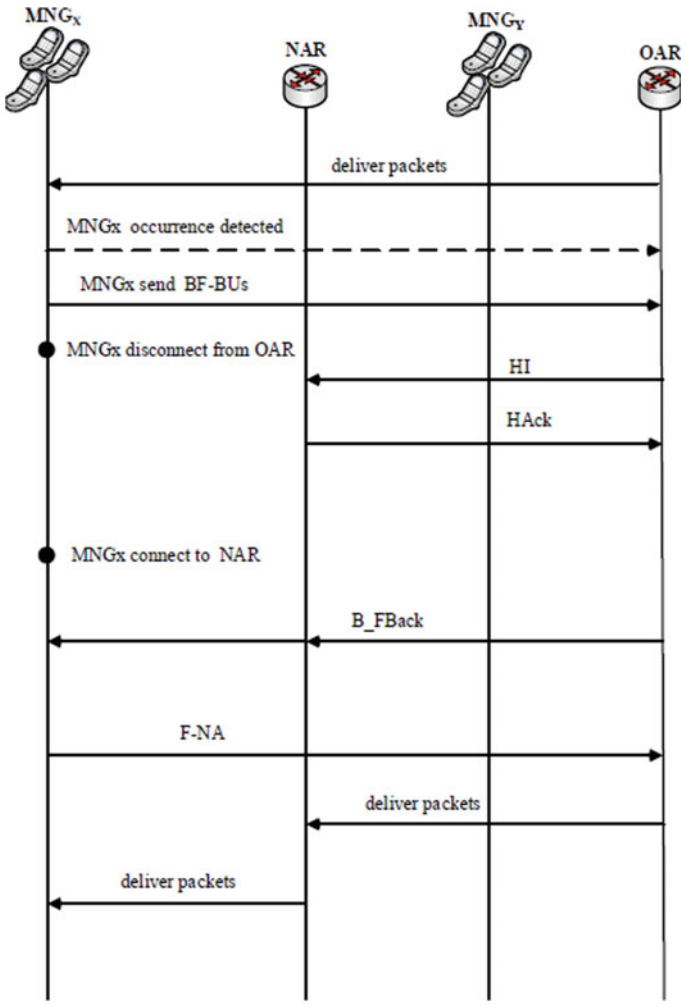


Fig. 5 The message flow diagram for EM-CSPMIPv6 scheme within FMIPv6 domain from CR-MN mechanism

In the CR-MN mechanism, we consider MNs which are receiving RSs at the same time from the same AR. In most of the protocols, the mobility related signaling are processed individually even when they simultaneously arrived to the same AR. This leads to increase the signaling cost and bandwidth waste. Subsequently, the proposed EM-CSPMIPv6 processes their registration simultaneously, as illustrated as follows.

Upon successfully detecting the movement of several MNs by the AR at simultaneously, the AR has two options to register the detected MNs. The first option, the related AR, upon detecting the movement of the MNs, sends a

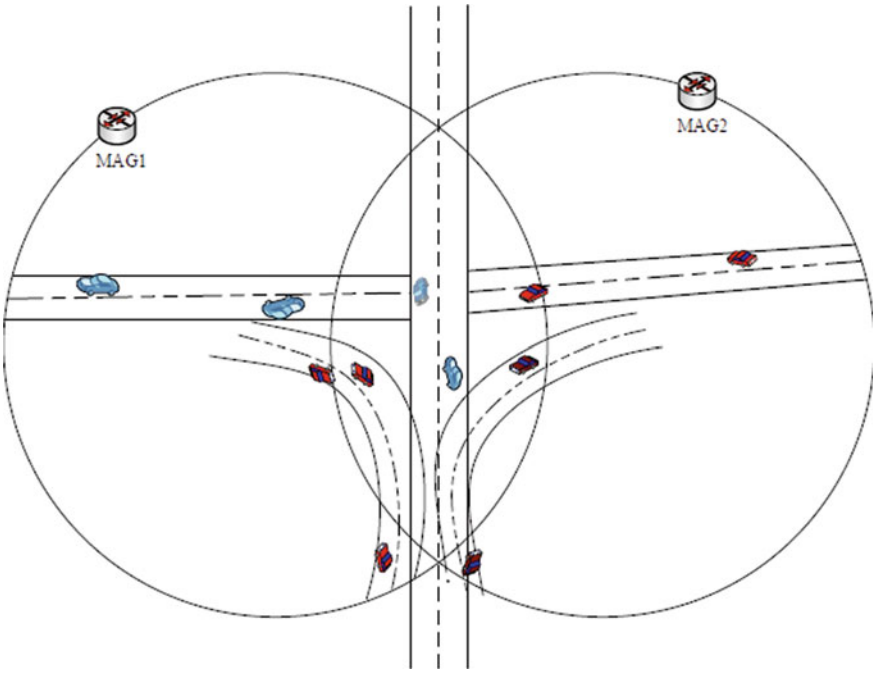


Fig. 6 Example of road diamond interchange

registration signaling control message including the detected MNs information. In the second option, a clock time is employed by the AR to collect the RS messages sent by the MNs which are intended to connect to AR. The time must be in milliseconds so that it does not affect the handover latency. In this work, we employed the prediction technique to avoid the delay that may happen from using the timer. The following steps explain the implementation scenario of the proposed mechanism.

The S_{th} threshold mentioned in Fig. 3 is used by the ARs to calculate the RSS. Upon detecting the MNs by the NAR and the RSS S_{th} threshold is exceeded, the AR discovers that the MNs are willing to move to another network. Consequently, the AR collects the addresses of movements' MNs and add them to a list to form a group of MNs. Then, the AR must wait till one of the grouped MNs, requests an actual handover. In this time, the AR sends one message containing the MAC addresses of the grouped MNs along with the MN that requested the actual handover to the Home Agent (HA) to preregister them.

We can clearly observe that the signaling cost in the CR-MN mechanism is dramatically reduced due to the processing all the mobility signaling simultaneously for several MNs. Furthermore, the burden on the bandwidth is reduced.

5 Performance Evaluation

To approve the superior effectiveness of our proposed EM-CSPMIPv6 scheme over the host-based, analyzing the proposed efficiency is performed to solve the bottlenecks associated with GB-FH method and FMIPv6.

5.1 Simulation Results

This section discusses the simulation results of the proposed scheme within host-based domains. NS2 is used to carry out intensive simulation experiments to evaluate the performance of our EM-CSPMIPv6 scheme with FMIPv6 protocol and GB-FH method. The FMIPv6 protocol and GB-FH method proposed in [8] are re-implemented and used as a benchmark for the EM-CSPMIPv6 scheme. Following that, the EM-CSPMIPv6 scheme is implemented in order to evaluate the handoff performance in terms of signaling cost and handoff latency. The overall signaling cost is the total number messages (in bites) exchanged between the MN, NAR, and PAR to register the MN. The experiments have been conducted based on two mechanisms explained in section four. The CN-MN mechanism manipulates the registration signaling of the MNs that are moving into groups, while the CR-MN one deals with the MNs that are initiating their mobility-related signaling at the same time.

Figure 7 shows the overall signaling cost during the handover initiation process. It can be clearly noticed that the FMIPv6 have the highest signaling cost in the CN-MN mechanism compared to GB-FH method and EM-CSPMIPv6 scheme. This is because the FMIPv6 protocol deals with each MN separately in the registration phase, which in turn increases the overall system signaling cost, especially when the number of moving MNs are very high. The high cost of signaling in the FMIPv6 protocol occurs as a result of using RS, RA, F-BU, F-Back, HI and HAck messages to register every new MN. On the other hand, GB-FH method outperforms the FMIPv6 protocol in terms of signaling cost. This can be attributed to the pre-registration that is performed by the MNG nodes to register multiple nodes in advance. Thus, messages such as RS and RA are eliminated from the registration phase during the MN registration. In contrast, the proposed EM-CSPMIPv6 shows the lowest signaling cost compared to GB-FH method and FMIPv6 protocol. This is due to several reasons such as MN registration time which is taken into account to reduce the false joining and the continuity of accepting the MNs joining requests, even when the HMN completes its registration phase, leading to increase the number of MNs pre-registration.

Figure 8 represents the handover latency time, which is the time between the last UDP packet received by the MN sent from OAR and first UDP packet received by the MN sent from NAR. As shown in Fig. 8, the handover latency in the standard FMIPv6 protocol grows with the increasing number of the MNs. The increase in the

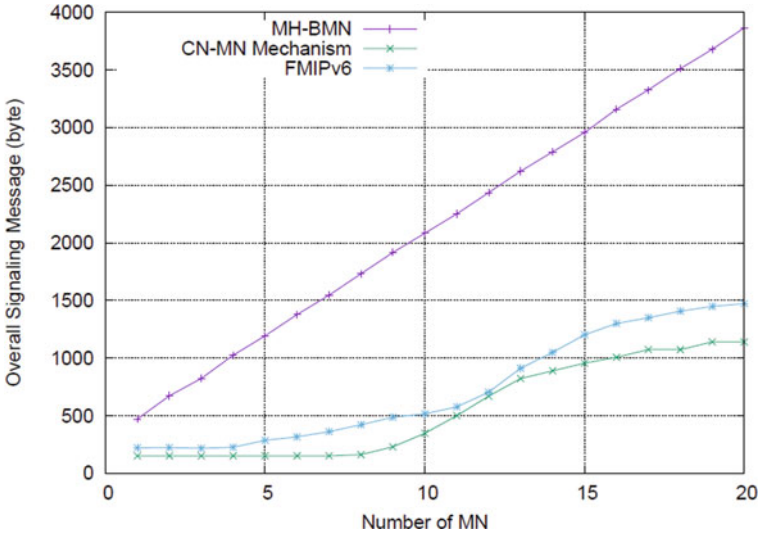


Fig. 7 Overall signaling load obtained from mechanism one

handover latency comes from the fact that every MN has its own handoff processing time. On the other hand, the GB-FH significantly reduces the handover latency due to the processing of the handoff for multiple MNs in advance. This leads to reduce the wireless channel contention, which in turn positively affects the system performance. In contrast, the proposed E-CSPMPv6 achieves the lowest handover delay compared to the FMIPv6 protocol and GB-FH method. The reason for that is, EM-CSPMIPv6 reduces the false prediction as well as increases the successful joining MNs through considering the first MN registration time and increasing the HMN accepting time.

In the CR-MN mechanism, Fig. 9 shows the results, which are obtained from applying EM-CSPMIPv6 according to the CR-MN mechanism steps. The EM-CSPMIPv6 in second mechanism has slightly overcome the GB-FH in terms of signaling overhead. The superiority is observed clearly when the number of MNs increases. This improvement occurs due to grouping and processing of MNs' mobility-related signaling for several MNs simultaneously. It can be observed from Fig. 9 that the proposed EM-CSPMIPv6 scheme has slight improvement compared to GB-FH method. This happens because of using a small number of MNs in our experiments. Moreover, the EM-CSPMIPv6 scheme in the CR-MN mechanism depends on the MNs which arrive simultaneously that may occur rarely. However, the EM-CSPMIPv6 scheme that is applied according to the CR-MN mechanism steps have a high signaling cost compared with the EM-CSPMIPv6 scheme from CN-MN mechanism. This comes because of inheriting the demerits of GB-FH such as false prediction by the EM-CSPMIPv6 scheme in the CR-MN mechanism. Moreover, FMIPv6 presents a poor performance due to performing handover signaling for each MN separately.

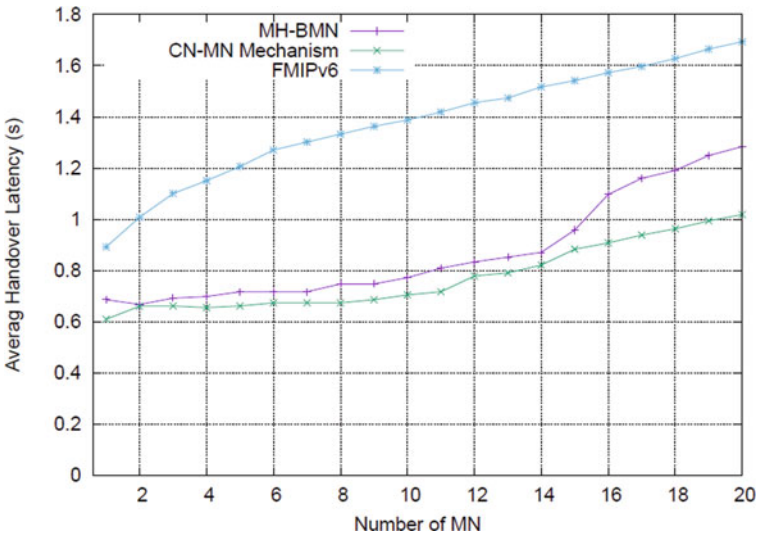


Fig. 8 Average handover latency obtained from mechanism one

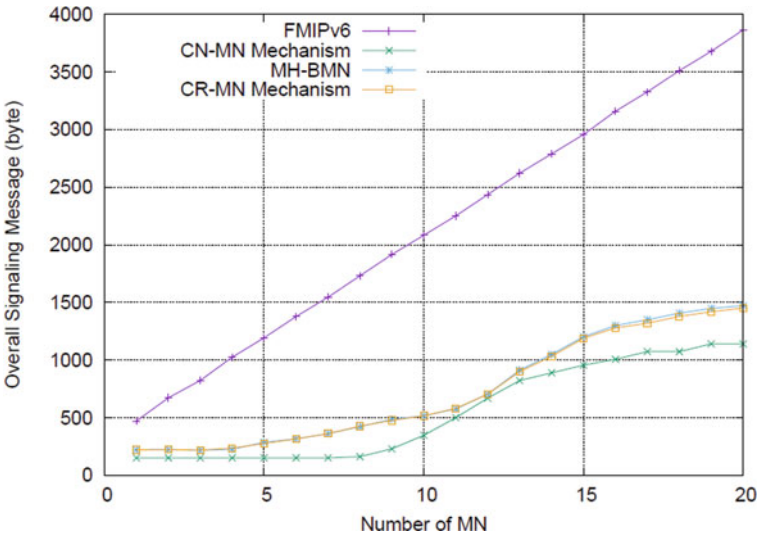


Fig. 9 Overall signaling load obtained from CR-MN mechanism

Figure 10 presents the effectiveness of E-CSPMUIPv6 when the number of MNs is high. It is obvious that the proposed EM-CSPMIPv6 scheme has the lowest signaling cost even when the number of MNs is increased. This superiority can be attributed to the fact that our mechanism performs one HI and HAck messages for a group of MNs. In contrast, the GB-FH method FMIPv6 protocol are greatly affected

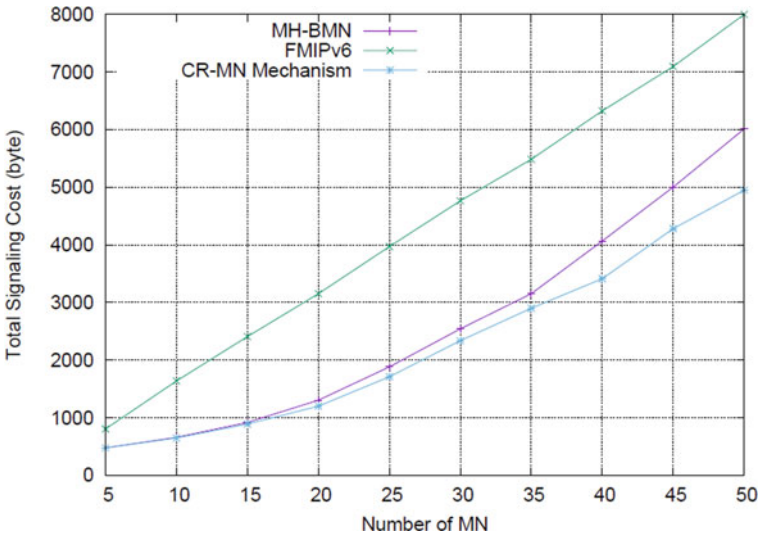


Fig. 10 Total signaling cost (byte)

by the increase in the MNs. This is because of performing a registration signaling for each MN separately in the FMIPv6 protocol and the false joining of MNs that occur due to applying GB-FH method.

6 Conclusion

An efficient EM-CSPMIPv6 scheme is implemented in this work to provide a handover without service disruption. This has been done through simultaneously manipulating handoff operations for group of MNs based on two suggested mechanisms.

CN-MN mechanism is implemented to perform Pre-registration and de-registration signaling for the connected and disconnected MNs at a closer distance from each other. The MN, which breaks the pre-defined threshold, must start clustering its neighboring MNs and must become the cluster head. Thereafter, the cluster head takes the responsibility to perform the initial binding registration processes in prior for its members. The mobility-related signaling messages like RA and RS have been extended to carry multiple addresses. The grouping of MNs, which are in motion into a cluster, serves in: low signaling, low handover latency and end-to-end delay.

The handover signaling are simultaneously addressed for the MNs, which perform link layer handover at the same time with the NAR. This is achieved via grouping the L2 requests of the MNs by the NAR and forwarding the L2 messages requests to the other FMIPv6 entities in one L2 message. The bandwidth burden is

lowered because of using number of served MNs during the handoff messages exchange.

Finally, simulation results demonstrate that the proposed EM-CSPMIPv6 scheme produces a better performance when is compared to FMIPv6 protocol and GB-FH method in terms of handover latency and signaling cost.

Acknowledgements This work is a part of a project submitted to Napier University.

Funding Our gratitude to UMT Grant with VOT 53,258: A GSM/GPS Tracker Technology in Tracking Monitoring of Beehives and VOT 53,290: Intelligent Mosquitoes Home System (IHMS) our financial support from.

References

1. Ghaleb M, Subramaniam S, Ghaleb SM (2019) An adaptive data gathering algorithm for minimum travel route planning in wsns based on rendezvous points. Symmetry 11(11):1326 [Online]. https://doi.org/10.3390/sym11111326
2. Man M, Bakar WA, Hwa LC, Yusoff WNJWM (2020) Dengueinnovation: a sustainability integrated approach for preventing and controlling of dengue diseases outbreaks via ir4. 0 technology. Int J 8(6)
3. Man M, Bakar WA, Razak MABBA (2020) Mus-tracker: an iotbased system in controlling and monitoring of beehives. Int J 8(6)
4. Johnson D (2004) IP mobility support for IPv6. RFC 3775
5. Koodli R (2009) Mobile IPv6 fast handovers. IETF, RFC 5568
6. Soliman H, Bellier L, Malki KE (2005) Hierarchical mobile IPv6 mobility management (HMIPv6). IETF, RFC 4140
7. Ghaleb SM, Subramaniam S, Zukarnain ZA, Muhammed A (2016) Mobilitymanagement for iot: a survey. EURASIP J Wirel Commun Netw 2016(1):165
8. Chiang M-S (2015) Fast handover control scheme for multi-node using the group-based approach. IET Netw 4:44–53(9), January 2015. [Online]. https://digital-library.theiet.org/content/journals. https://doi.org/10.1049/iet-net.2013.0150
9. Akyildiz IF, Xie J, Mohanty S (2004) A survey of mobility management innext-generation all-IP-based wireless systems. IEEE Wirel Commun 11(4):16–28
10. Hong S, Kim D, Ha M, Bae S, Park SJ, Jung W, Kim J (2010) SNAIL: an IP-based wireless sensor network approach to the internet of things. IEEE Wirel Commun 17(6):34–42
11. Ma W, Fang Y (2004) Dynamic hierarchical mobility management strategy for mobile IP networks. IEEE J Sel Areas Commun 22(4):664–676
12. Mohanty S, Akyildiz IF (2007) Performance analysis of handoff techniques based on mobile IP, TCP-Migrate, and SIP. IEEE Trans Mob Comput 6(7):731–747
13. Chen WE, Li SH (2013) Client-based Internet Protocol version 4-Internet Protocol version 6 translation mechanism for Session Initiation Protocol multimedia services in next generation networks. IET Netw 2(3):115–123
14. Yen Y-S, Chen L-Y, Chi T-Y, Chao H-C (2013) A novel predictive scheduling handover on mobile IPv6. Telecommun Syste 52(2):461–473 [Online]. https://doi.org/10.1007/s11235-011-9448-x
15. Ramani I, Savage S 92005) SyncScan: practical fast handoff for 802.11 infrastructure networks. In: Proceedings IEEE 24th annual joint conference of the IEEE computer and communications societies, vol 1, March 2005, pp 675–684

16. Song Q, Jamalipour A (2006) A time-adaptive vertical handoff decision scheme in wireless overlay networks. In: 2006 IEEE 17th international symposium on personal, indoor and mobile radio communications, Sept 2006, pp 1–5
17. Chen X, Qiao D (2010) HaND: fast handoff with null dwell time for IEEE 802.11 Networks. In: 2010 Proceedings IEEE INFOCOM, March 2010, pp 1–9
18. Soliman H, Castelluccia C, Elmalki K, Bellier L (2008) Hierarchical mobile IPv6 (HMIPv6) mobility management. RFC 5380, October, October 2008 [Online]. http://www.rfc-editor.org/info/rfc5380
19. Jaron A, Pangalos P, Mihailovic A, Aghvami A (2012) Proactive autonomicload uniformisation with mobility management for wireless internet protocol (ip) access networks. IET Netw 1:229–238(9) [Online]. http://digital-library.theiet.org/content/journals/, https://doi.org/10.1049/iet-net.2011.0009
20. Petrescu A, Wakikawa R, Thubert P, Devarapalli V (2005) Network mobility (NEMO) basic support protocol. IETF RFC. 4063
21. Kim JH, Hong CS, Shon T (2008) A Lightweight NEMO protocol to support6LoWPAN. ETRI J 30(5):685–695
22. Kim JH, Hong CS, Okamura K (2007) A routing scheme for supporting network mobility of sensor network based on 6LoWPAN. Springer, Berlin,Heidelberg, pp 155–164. [Online]. https://doi.org/10.1007/978-3-540-75476-3
23. Chai R, Zhao YL, Chen QB, Tang L (2009) Group mobility in wireless sensor network. In: 2009 international conference on wireless communications signal processing, Nov 2009, pp 1–5
24. Chai R, Zhao YL, Chen QB, Dong T, Zhou WG (2010) Group mobility in6LoWPAN-based WSN. In: 2010 international conference on wireless communications signal processing (WCSP), Oct 2010, pp 1–5
25. Ghaleb SM, Subramaniam S, Ghaleb M, Ejmaa ME et al (2019) An efficient group-based control signalling within proxy mobile ipv6 protocol. Computers 8(4):75
26. Ghaleb SM, Subramaniam S, Zukarnain ZA, Muhammed A (2018) Load balancing mechanism for clustered pmipv6 protocol. EURASIP J Wirel Commun Netw 2018(1):1–23
27. Huang C-M, Dao D-T, Chiang M-S (2018) A bursty multi-node handover scheme for mobile internet using the partially distributed mobility management (bmh–dmm) architecture. Telecommun Syst 69(1):113–130

Design and Simulation a Patch Antenna for Dual-Band Frequencies

Mohammed Salah Abood, Ahmed Shamil Mustafa, Maha Fathy, Mustafa Maad Hamdi, and Mustafa Sabah Noori

Abstract In communication systems, wireless technology is the most thoroughly studied environment and communication schemes were not practiced without understatement of the antennas' operation. The proposed antenna can be used for Long Term Evolution (LTE), with strong coverage and improved power, for the indoor and outdoor regions. The antenna has Omni-directional pattern at lower and upper operation at frequency LTE band 7 (2.6 GHz) and band 3 (1.8 GHz). The concept antenna was produced with the substratum FR-4. The parameters of dispersion matrices, radiation pattern, gain, directive, VSWR, return loss can be calculated. The prototypes and calculations were done with Microwave Studio (CST). The shown array antenna can be extended in particular to mobile and small base stations.

Keywords LTE · CST · Dual-band · Patch antenna

1 Introduction

The rapid proliferation of interactive devices and emerging digital infrastructure growth and the increased usage of a mobile terminal have driven the research into modern network architectures and wireless world requirements [1]. Moreover, the creation of modern and exciting wireless service has been accelerating by rising processing capacity, memory and high-end graphics functionality. Personal video

M. S. Abood (✉) · M. Fathy
School of Information and Electronics Engineering, Beijing Institute of Technology, Beijing, China

A. S. Mustafa · M. M. Hamdi
Department of Computer Engineering Techniques, Al-Maarif University College, Al-Anbar, Iraq

M. S. Noori
Department of Computer and Communication Systems Engineering, Universiti Putra Malaysia, Selangor, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_35

recorders, on request recordings, various programming, interactivity, mobile telephony and media broadcasting have allowed audiences to personalize their viewing and to communicate their desires to broadcasters [2–4]. Wireless networking infrastructure is a new need and has quickly expanded with fresh concepts, where it begins with the first LTE and LTE-Advanced technology (LTE-A). LTE or LTE-A was the high-speed broadband network for cell phones and data terminals based on the Global Mobile Infrastructure (GSM). Global Cell Broadband has been project to amount mobile user growth over the next 3 years by an estimate of 50% each year [5, 6]. In wireless technology, especially in mobile devices, data traffic and signal coverage will be quite significant. Therefore, the Femtocell Network has been developed by the 3rd generation partnership project (3GPP) to support data traffic and signal coverage from macrocell base stations [7, 8].

LTE technology has potential peak data speeds for the downlink and uplink of 300 Mbps and 75 Mbps respectively. LTE technology has been used for the first time in Europe, and the 1.8 and 2.6 GHz frequency bands have been assigned. The 800–960 MHz and 3.4–3.8 GHz bands are now allowed to run LTE bands as the need for fast data rates grows [9–11]. The antenna developments are also an immense obstacle for them to support the high demand for these service classes. In recent years, a massive number of antennas have been built and produced to cover these LTE frequency bands which can only provide single and dual bands including single [12, 13] and double [14] bands.

Patch antennas perform a critical part in the wireless networking device environment in recent days. Microstrip antennas are very charming because of their low profile, their low weight and their superficiality. A significant number of microstrip antennas have been produced to be used in wireless applications [15–18]. Miniature antenna scale, broad band and multi band antennas are expected to fulfill potential advancement of wireless networking technologies. The microstrip antenna is going to be a high-quality rival for several connectivity products. Antenna contains both dielectric layer and ground plane on the other hand. The fine feature of microstrip patch antenna, such as small weight, stubby body planar configuration, cheap fabrication costs and the capacity to integrate with microwave integrated circuits technologies, make the patch antennas suitable for wireless networking applications, cellular phones, pagers, radar, and satellite communications equipment items.

The antenna's size is decreased effectively by cutting a slot on the microstrip road. The goal of this paper is to develop and simulate a dual-band microstrip patch antenna using a CST microwave studio at frequency LTE band 7 (2.6 GHz) and band 3 (1.8 GHz) on the FR4-epoxy substratum of 1.6 mm substrate height (h) = 1.6 mm and conductor height (t) = 0.035 mm. This antenna is design to function within the spectrum of frequencies between the two (1.8–2.6 GHz).

2 Related Work

Due to their versatility and its diverse uses, wireless networks have exciting research fields. However, there are also a variety of problems to be researched and analyzed. This unit explains the use of a patch antenna for different wireless applications.

In the study [19], a basic microstrip patch antenna with a resonant frequency of 2.4 GHz is constructed in the CST microwave studio. The benefit of the built antenna is 8.27 dB and 1.18 VSWR.

In [20], the patch antennas scheme and emulation are used for cell users. The return loss and the numerous benefit plots were analyzed along with the radiation style.

In the research [21], the antenna targeted at 2.4 GHz frequency resonance operation for wireless local area network (WLAN). Different shapes of an antenna exist in this design and this antenna is designed to maintain the above resonant frequency. Besides, this antenna is applied to the dielectric substratum of FR 4 Epoxy and this configuration has been tested according to the return loss and VSWR.

In work [22] proposed Π-Shaped Slot Dual Band Antenna, the antenna return losses are –16.2 dB and –12.94 dB, significantly less for 2X2 MIMO pi antenna. Gain values have been obtained, which satisfy the WLAN and Wi-Max applications. The antenna gain values are not very high.

The authors' proposed [23] is a dual-band dual-polarized planar antenna with sharp flat-top cut off radiation for the base stations. The range comprises of eight to eight antenna components organized with standardized spacing and is allocated with tailored uniform enthusiasm to achieve flat top and sharp cut off of low side lobes radiation patterns.

3 Antenna Parameters

This segment analyzes the different parameters such as VSWR, Return Loss, Antenna Gain, Directivity, Antenna Performance and Bandwidth [24].

- Gain: Antenna gain is characterized as a ratio of energy, in a given direction, to the intensity of radiation that would be received if the antenna power was isotopically radiated. G = 4 μ.U(first, next)/pin is the formula for the benefit, where, U(first), side), the pin is the input power.
- Radiation pattern: The radiation pattern is known as a mathematical feature or a graphical representation of the antenna's radiation characteristics according to its space co-ordinates.
- Antenna efficiency: It's a ratio of overall antenna output to an antenna's input power.
- VSWR: The VSWR = Vmax/Vmin voltage standing wave ratio is established.

- It could range from 1 to 2.
- Return loss: Return loss reflects signal strength as a system is embedded in a transmission cable. Therefore, the RL is a parameter close to the VSWR to demonstrate how well the transmitter and antenna matched. The RL is as follows: RL = –20 log10 (mode) dB for a perfect fit between transmitter and antenna = 0 and RL = oscillation of power, indicating that no control will be returned, whereas the value of an RL = 1 is RL = 0 dB, suggesting the representation of all power in the event. A VSWR 2 is permissible for practical applications since it meets an RL of –9.54 dB.

4 Proposed Antenna and Parameters

Figure 1 shows the geometry of the front view and architecture of the planned CST Microwave Studio software patch line antenna with dual-band operation. The antenna feed point measurements and location was intended to achieve the optimum antenna impedance. The following criteria are used for the design of the proposed antenna. A number of parameters have been evaluated in this report using CST Microwave Studio methods.

A rectangular microstrip antenna is designed for a dual-frequency UWB device communication use. The suggested rectangular patch antenna has been conceived with dielectric substrate Fr4 = 4.3, substrate height (h) = 1.6 mm and conductor height (t) = 0.035 mm. This antenna operates at a frequency of 1.8 and 2.6 GHz. The other parameters for the antenna are the patches' width (W) = 51 mm, the patches' length (L) = 39.4 mm, the ground plane's width and length, and the support (Wg) = 102 with (Lg) = 78.8, the patches' width and the microstrip field line (Gpt) = 1 mm, and the patch patching antenna and Table 1 are shown as description parameters for the antenna. Figure 2 show the rectangular patch antenna.

This research was carried on the philosophy, operating concept and characteristics of this antenna. There are several microstrip antenna research methods, so we

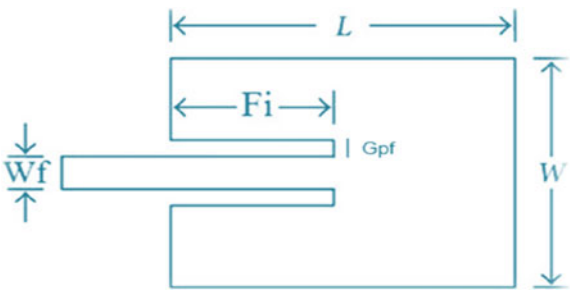


Fig. 1 Geometry of the antenna designed

Table 1 Parameters values

Parameters	Mm
W	51
L	39.4
FI	12.0075
Wf	3.1
Lg	2*L
Gpf	1
Wg	2*W
Ht	0.035
Hs	1.6

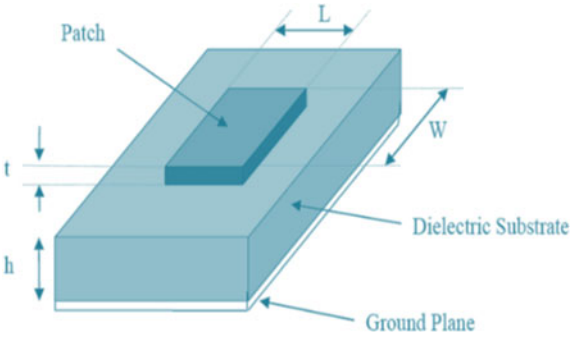


Fig. 2 Rectangular patch antenna

can conveniently select the transmission line mode list. The equations of the transmission line as follows:

Step 1: At the first, the C = 299,792,458, $f_{\circ}$ = 1.8 with 2.4 GHz, and Ɛr = 4.3. To find Width (W) we used Eq. (1).

$$W = \frac{c}{2f^{\circ}\sqrt{\frac{Er+1}{2}}} \tag{1}$$

Step 2: Eq. (2) establishes the efficient dielectric constant:

$$\varepsilon_{\text{reff}} = \frac{\varepsilon_{\text{r}}+1}{2} + \frac{\varepsilon_{\text{r}}-1}{2}\left[1+12\left(\frac{\text{h}}{\text{w}}\right)\right]^{-\frac{1}{2}} \tag{2}$$

Step 3: The sufficient length calculated from Eq. (3):

$$\text{L}_{\text{eff}} = \frac{\text{C}}{2\text{f}_{\text{r}}\sqrt{\varepsilon_{\text{reff}}}} \tag{3}$$

Step 4: The fringing length (ΔL) from Eq. (4):

$$\Delta L = 0.1412h\frac{(\varepsilon_{reff}+0.3)\left(\frac{W}{h}+0.264\right)}{(\varepsilon_{reff}-0.258)\left(\frac{W}{h}+0.8\right)} \tag{4}$$

Step 5: The actual length Land the width and length of the Ground calculated by used Eq. (5):

$$L = L_{eff} - 2\Delta L \tag{5}$$

Step 6: The microstrip feed line (inset-fed) with input impedance 50 Ω and Gpf 1 mm by Eq. (6):

$$\begin{aligned} fi = 10^4\big(&0.001699*\varepsilon_r^7 + 0.13761*\varepsilon_r^6 - 6.1783*\varepsilon_r^5 + 93.187*\varepsilon_r^4 \\ &-682.69*\varepsilon_r^3 + 2561.9*\varepsilon_r^2 - 4043*\varepsilon_r + 6697\big)*\frac{L}{2} \end{aligned} \tag{6}$$

Step 7: The width of Microstrip width from Eq. (7):

$$Zc = \frac{120\pi}{\sqrt{\varepsilon_{eff}}\left[\frac{Wf}{h}+1.393+0.667*\ln\left(\frac{Wf}{h}+1.444\right)\right]} \tag{7}$$

5 Simulated Results

Patch antenna efficiency in Fig. 3 is measured in terms of coefficient of reflection (S11), radiation properties and gain. Figure 4 displays S11's simulated result. The antenna shows a dual-band operating frequency at 1.8 GHz and 2.6 GHz respectively –15.327 dB and –18.724 dB. The simulated frequency ranges of 1.77–1.83 GHz and 2.55–2.65 GHz are acceptable.

The simulated patch antenna gave a resonant frequency of 1.8 GHz. The simulated return loss is found to be –15.327 dB. We get return loss peaks of the dual-band patch antenna at 1.8 GHz, and 2.6 GHz are—15.327 dB and –18.724 dB respectively. The radiation pattern at both frequencies in polar form is shown in the figure above. Figures 5, 6, 7 and 8 shows the excitation signal of the proposed antenna, the excitation signal S-Parameter, the Fairfield Radiation (Abs) pattern when the frequency is 1.8 GHz, and Fairfield Radiation (Abs) pattern when the frequency 2.6 GHz is present. Based on our results and when compare it with work which done by [25], the dual band recorded a batter value in dB.

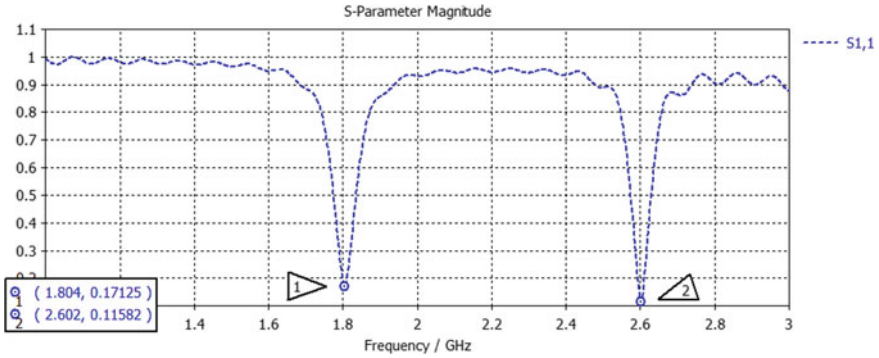


Fig. 3 S-parameter magnitude

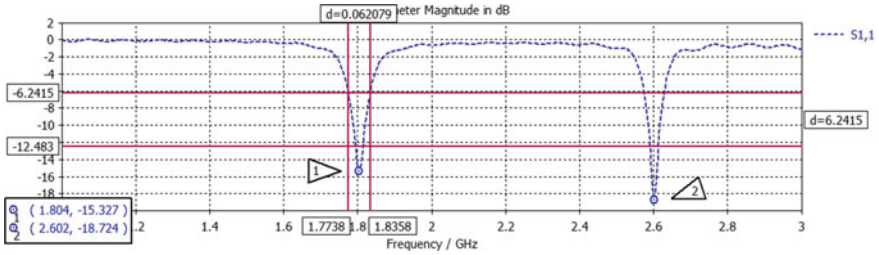


Fig. 4 S11 simulated result with bandwidth 62 MHz

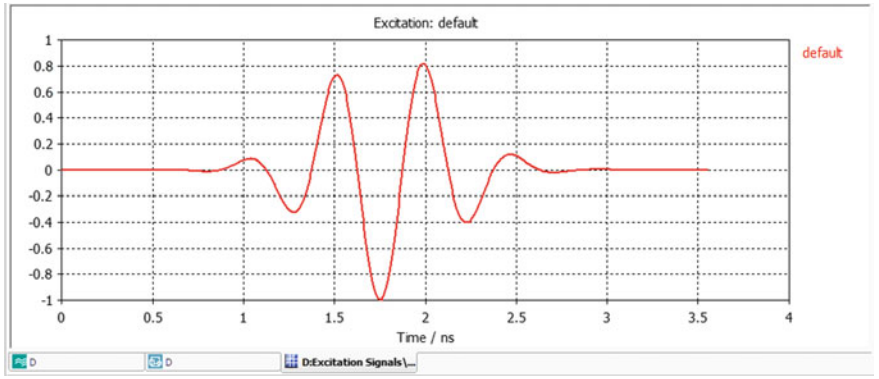


Fig. 5 Excitation signal of the proposed antenna

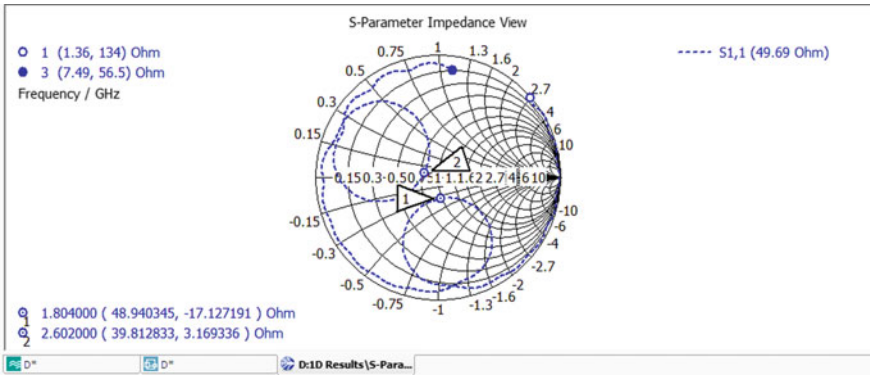


Fig. 6 The excitation signal S-Parameter of the proposed antenna

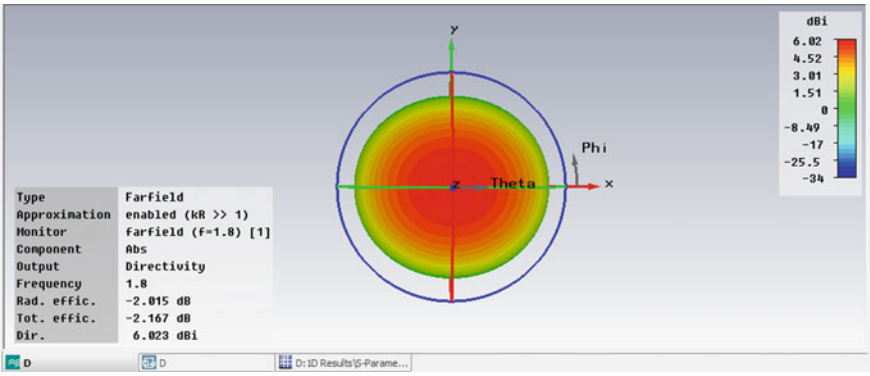


Fig. 7 Fairfield Radiation (Abs) pattern when the frequency is 1.8 GHz

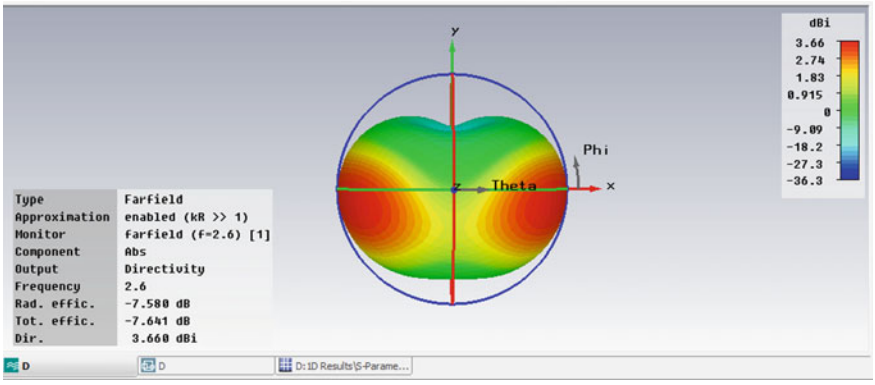


Fig. 8 Fairfield Radiation (Abs) pattern when the frequency is 2.6 GHz

6 Conclusions

Dual-band creation and dual-polarized LTE Patch antenna are present. The planned antenna works are—15.327 dB and –18.724 dB respectively return losses spanning Wi-Max bands under the linearly polarized 1.8–2.6 GHz range. Robust radiation pattern findings have been produced that seems to be sufficient for the applications envisaged. The antenna then suggested displays dual-polarized operational activity on LTE bands. Present LTE band antenna distributions are in linear directions. Furthermore, the antenna features good directional radiation patterns, good matching impedance and high gains, making it a right candidate for LTE communication applications with three-band. The more findings are discussed in depth in our paper. For future work, this antenna may be developed in the laboratory. Further research may be carried out to calculate and examine the function antenna, such as radiation pattern, bandwidth, and return loss. This is essential for the antenna to operate in the required features. To test the fabricated antenna, the Vector Network Analyzer (VNA) may be used to show the real return failure, the S11 and the bandwidth relative to the simulation data. This is necessary to ensure that the antenna functions according to the desired features.

References

1. Huang H (2013) Flexible wireless antenna sensor: a review. IEEE Sens J 13(10):3865–3872
2. Mustafa AS, Abdulelah AJ, Ahmed AK (2019) Multimodal biometric system Iris and fingerprint recognition based on fusion technique
3. Hamdi M (2018) Hybrid tow feature extraction descriptor for shape pattern recognition. Aust J Basic Appl Sci
4. Abood MS, Ismail M, Nordin R (2016) A quiz management system based on P2P near-field communication on Android platform for smart class environments. In: 2016 international conference on advances in electrical, electronic and systems engineering (ICAEES), pp 83–88. IEEE
5. Ahmed AU, Islam M, Azim R, Ismail M, Mansor MF (2014) Microstrip antenna design for femtocell coverage optimization. Int J Antennas Propag 2014
6. Abood MS, Thajeel IK, Alsaedi EM, Hamdi MM, Mustafa AS, Rashid SA (2020) Fuzzy logic controller to control the position of a mobile robot that follows a track on the floor. In: 2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT), pp 1–7. IEEE
7. Ibrahim A, Ibrahim M, Dewan R (2017) Dual band patch antenna for long term evolution (LTE) application in femtocell network. In: 2017 7th IEEE international conference on control system, computing and engineering (ICCSCE), pp 310–313. IEEE
8. Mustafa AS, Hamdi MM, Mahdi HF, Abood MS (2020) VANET: towards security issues review. In: 2020 IEEE 5th international symposium on telecommunication technologies (ISTT), pp 151–156. IEEE
9. Yu K, Li Y, Luo X, Liu X (2016) A modified E-shaped triple-band patch antenna for LTE communication applications. In: 2016 IEEE international symposium on antennas and propagation (APSURSI), pp 295–296. IEEE

10. Hamdi MM, Audah L, Rashid SA, Al-Mashhadani MA (2020) Coarse WDM in metropolitan networks: challenges, standards, applications, and future role. In: 2020 Journal of physics: conference series, vol 1530(1). IOP Publishing, p 012062
11. Hamdi M (2018) BER Vs Eb/N0 BPSK modulation over different types of channel. Aust J Basic Appl Sci
12. Redzwan F, Ali M, Tan MM, Miswadi N (2014) Dual-band Planar Inverted F Antenna with parasitic element for LTE and WiMAX mobile communication. In: 2014 international symposium on technology management and emerging technologies, pp 62–67. IEEE
13. Chellappan K, Mustafa AS, Mohammed MJ, Thajeel AM (2015) Layered defense approach: towards total network security. Int J Comput Sci Bus Inf 15(1)
14. Dioum I, Diallo A, Farssi SM, Luxey C (2014) A novel compact dual-band LTE antenna-system for MIMO operation. IEEE Trans Antennas Propag 62(4):2291–2296
15. Gupta R (2009) Printed tri-band monopole antenna structures for wireless applications. Microw Opt Technol Lett 51(7):1781–1785
16. Hamdi MM, Mustafa AS, Mahd HF, Abood MS, Kumar C, Al-shareeda MA (2020) Performance analysis of QoS in MANET based on IEEE 802.11 b. In: 2020 IEEE international conference for innovation in technology (INOCON), pp 1–5. IEEE
17. Mahdi HF, Hamdi MM, Abood MS, Abdulelah AJ, Mohammed RQ (2020) An overview on binary and non-binary codes and challenges in low parity-check-density. In: 2020 4th international symposium on multidisciplinary studies and innovative technologies (ISMSIT), pp 1–6. IEEE
18. Jasim NA, Abbood S, Salah M (2018) Speech synthesis using neural network. Open Sci J 3(1)
19. Afridi MA (2015) Microstrip patch antenna—designing at 2.4 GHz frequency. Biol Chem Res 2015:128–132
20. Bhalla D, Bansal K (2013) Design of a rectangular microstrip patch antenna using inset feed technique. IOSR J Electron Commun Eng 7(4):08–13
21. Adegoke OM, Eltoum IS (2014) Analysis and design of rectangular microstrip patch antenna at 2.4 GHz WLAN applications. Int J Eng Res Technol (IJERT) 3(8)
22. Ravindiran A, Jeyalakshmi V (2017) Design and implementation of Π-shaped slot dual band antenna for WLAN/WiMAX applications. In: 2017 International conference on nextgen electronic technologies: silicon to software (ICNETS2), pp 75–78. IEEE
23. Wang C, Chen Y, Yang S (2018) Dual-band dual-polarized antenna array with flat-top and sharp cutoff radiation patterns for 2G/3G/LTE cellular bands. IEEE Trans Antennas Propag 66 (11):5907–5917
24. Parmar N, Saxena M, Nayak K (2014) Review of Microstrip patch antenna for WLAN and Wimax application. Int J Eng Res Appl 4(1):168–171
25. Al-Hiti AS (2019) Design of rectangular microstrip patch antenna for WLAN and WiMAX applications. ARPN J Eng Appl Sci 14(2)

Centroidal Polygon: A New Modified Euler to Improve Accuracy of Resistor-Inductor (RL) Circuit Equation

S. Nooraida, Z. Nur Shahirah, S. Noor Suhana, and N. Akhyari

Abstract The Euler method provides a simple procedure for solving Ordinary Differential Equations (ODE). Resistor-inductor (RL) circuit equation is one of the circuit equations applicable to the first order differential equation. Although the Euler method gives a simple solution for solving ODE, questions have been raised on the accuracy, complexity and processing time that compelled the researchers to use more sophisticated methods such as Runge Kutta, Adams–Bashforth etc. To ensure that numerical method is applicable for Electrical mathematics, the modified Euler method will be adopted in each RL circuit equation. This research uses the Centroidal Polygon (CP) method, which is a subset to the modified Euler method, to increase accuracy for the solution. To ensure that the CP method is numerically computable with sufficient accuracy for applications outside mathematical theorem, the results depended on the maximum error of the simulation. In addition to that, the accuracy of the circuit equation depends on the time constant (t). This research utilized Scilab 6.0 software to generate and analyze the results required to examine the accuracy of the simulation. The results were compared to the previous scheme, Polygon (P) and Harmonic Polygon (HP). The results found that the CP scheme yielded better accuracy and feasibly an alternative scheme.

Keywords Resistor Inductor (RL) circuit · Centroidal mean · Accuracy

S. Nooraida (✉) · Z. Nur Shahirah · S. Noor Suhana · N. Akhyari
University College TATI, Teluk Kalong, Kemaman, Terengganu, Malaysia
e-mail: nooraida@uctati.edu.my

Z. Nur Shahirah
e-mail: shahirah.zulkifli.md@gmail.com

S. Noor Suhana
e-mail: suhana@uctati.edu.my

N. Akhyari
e-mail: akhyari@uctati.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_36

1 Introduction

The Euler method is one of the oldest methods to solve mathematical problems. The Euler method was named after Leonhard Euler, who wrote a book on integral calculation. The book was written from 1768 to 1870 [1]. The Euler method returned accurate values but by only using a small step size [2]. However, a small step size leads to slow computation. Therefore, researchers opted for different approaches that provided more precise solutions with faster computation, such as Runge–Kutta and Adam Bashforth, as compared to the Euler form. However, the approaches are far more complex in comparison to the Euler method [3].

Euler's Method is a computational approach to solve problems with initial values. The method from Euler is based on the insight of differential equations. The tangent line is just a decent approximation over a short interval as shown by the researchers in their previous study. Therefore, studies were conducted by the researchers to create a new tangent for improving the accuracy of Euler's method [4]. The improved Euler method is mainly applied to mathematical areas. After it has been proven to solve mathematical problems, the researchers will then test the solution in other areas. This is to ensure that the Euler method can be extended to other fields, specifically in the field of engineering. The Euler method also proved its application in mechanical engineering, particularly in computational fluid dynamics (CFD), to study the behaviour of fluid flow [5, 6]. Hence, in this study, we will explore the application of Euler's method in electrical engineering. The targeted area in electrical is in circuit equations.

In electrical engineering, there are two conditions of electric circuit equation that are applicable to the first order equation namely, Resistor–Capacitor (RC) and Resistor-Inductor (RL) [7]. RC circuits consist of a resistor and a capacitor. Meanwhile RL circuits consist of a resistor and an inductor. This research will concentrate on the equation of the RL circuit equation for the analysis.

This analysis will be performed by comparing the proposed scheme against Polygon (P) and Harmonic Polygon (HP). The researcher will conduct an analysis of accuracy between the proposed scheme against Polygon and Harmonic Polygon. Scilab 6.0 software [8] is used to model the RL circuit equation and subsequently solving the equation. There were three different RL circuit equations used for testing, to ensure that the Centroidal Polygon scheme method can be applied into any of the value problems.

Centroidal Polygon (CP) scheme is a combination of the enhanced Euler's method and the mean concept. Centroidal mean is used to get better accuracy when incorporated with the enhanced Euler's method. CP scheme was tested on RL circuit equations to discover new methods that can be employed in electrical application. The differential step size, h, of 0.1, 0.01, and 0.001 were used in testing the schemes. The main aim of the study is to propose a new and enhanced Euler's method called CP scheme and improve its accuracy. Besides that, this research was also aimed to study the speed of the CP scheme in solving first order Resistor-Inductor (RL) circuit equations.

1.1 Enhanced Euler's Method

Euler's scheme provides benefits such as easy execution and low computing costs. Euler's schemes can deliver a sophisticated solution if the step size, *h,* of the measurement is minimized. However, shrinking the step size may affect the instability of Euler's scheme. Large step size *h* selection is also expected to create an inaccurate solution in solving ODE problems. Several recent studies were conducted to improve the original Euler's scheme to complete the first stage of ODE such as [9–11]. They used the modified Euler's concept in their study. Enhanced Euler's method is an improvement to the original Euler's method.

In the modified Euler's method, the average of two points on the slope of x_i and x_{i+1} is taken. This average value is then applied to y_{i+1}. In [8] which creates an alternative solution, that is able to match the accuracy of performance with the Euler's method and become a more stable scheme. The scheme obtained an exact solution by using the original Euler's concept and mean. For this study, the focus was on improving the performance of the modified Euler's scheme using the concept of centroidal mean.

1.2 Resistor-Inductor (RL) Circuit Equation

RL circuit is an electric circuit composed of resistors and inductors driven by a voltage or current source. A first-order RL circuit is composed of one resistor and one inductor, which is the simplest type of RL circuit. Figure 1 below shows a simple first-order RL circuit with a switch. When the circuit is open, the voltage across the resistive load will be at its maximum and when the switch is closed, the voltage from the battery is shared between the inductor and the resistive load. The modified Euler's scheme returns a small error in every calculation process in the electrical circuit equation [12].

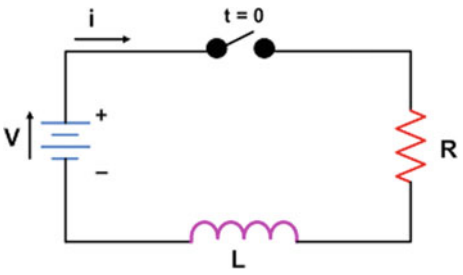


Fig. 1 RL circuit with a switch

2 Methodology

2.1 Construction of Centroidal-Polygon Scheme

Figure 2 shows how the proposed scheme is developed. This scheme is developed by combining Euler (R) and Mean (T). The original Euler's scheme, with the general formula, $n + 1 = y_n + hf(x_n, y_n)$, was selected as the basis for developing the scheme. Centroidal mean (T1), was selected to develop this proposed scheme. This combination of Euler (R) and mean (T1) produces the proposed scheme known as Centroidal Polygon (CP) (RT_1).

The Euler's method is $y_{n+1} = y_n + \Delta t f(x_n, y_n)$, which is used initially to compute $f(x_0, y_0)$. By doing this step, the slope of the line that is tangent to the solution will curve at the point (x,y). Recall that the slope is defined as the change in y divided by the change in t. It can be presented as $\Delta y/\Delta t$.

$$y_{n+1} = y_n + \Delta t f(x_n, y_n) \tag{1}$$

CP scheme on the centroidal mean equation will be applied into the function of $f(x, y)$,which is in the Euler's equation. CP scheme which was implied in Eq. (1) improves the Euler's equation based on Eqs. (2) and (3).

The improved Euler uses the mean concept. The mean used is centroidal mean, where the point is considered as the centre of two points, as written in Eqs. (2) and (3).

$$\frac{2x_0^2 + x_0x_1 + x_1^2}{3x_0 + x_1}, \frac{2y_0^2 + y_0y_1 + y_1^2}{3y_0 + y_1} \tag{2}$$

$$\frac{2x_0^2 + x_0(x_0 + h) + (x_0 + h)^2}{3x_0 + (x_0 + h)}, \frac{2y_0^2 + y_0(y_0 + h) + (y_0 + h)^2}{3y_0 + (y_0 + h)} \tag{3}$$

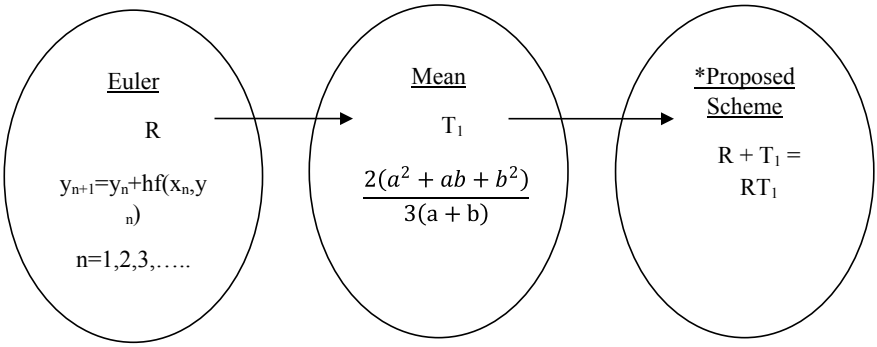


Fig. 2 Proposed scheme

To improve the equation, the CP equation in Eq. (3) will be implemented into the Eq. (1). The new equations formed are as in Eq. (4)

$$\frac{y-y_0}{h}=f\left(\frac{2x_0^2+x_0(x_0+h)+(x_0+h)^2}{3x_0+(x_0+h)},\frac{2y_0^2+y_0(y_0+h)+(y_0+h)^2}{3y_0+(y_0+h)}\right)$$

$$y-y_0=hf\left(\frac{2x_0^2+x_0(x_0+h)+(x_0+h)^2}{3x_0+(x_0+h)},\frac{2y_0^2+y_0(y_0+h)+(y_0+h)^2}{3y_0+(y_0+h)}\right)$$

$$y=y_0+hf\left(\frac{2x_0^2+x_0(x_0+h)+(x_0+h)^2}{3x_0+(x_0+h)},\frac{2y_0^2+y_0(y_0+h)+(y_0+h)^2}{3y_0+(y_0+h)}\right) \quad (4)$$

Euler can be more accurate and simulated faster by modifying the equation using the midpoints as shown in Eq. (4).

2.2 Experimental Setup

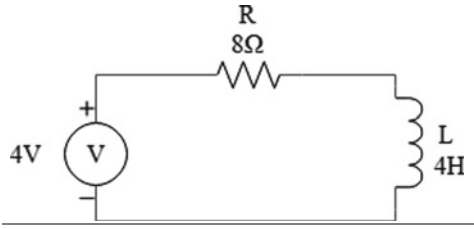
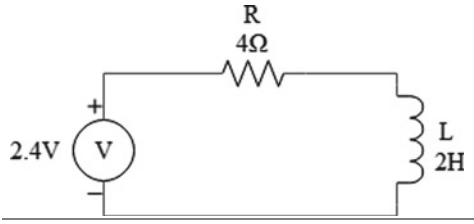
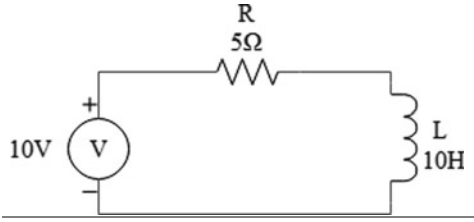
This experiment will be conducted by testing RL circuit equation on Polygon, Harmonic Polygon and Centroidal Polygon scheme. All the schemes will use three different voltage values. Before the schemes are tested into the RL circuit equation, the schemes are computed into the Ordinary Differential Equation (ODE), to ensure that the schemes are computable. The RL circuit equation is a linear equation. Before testing the schemes into the RL circuit equation, the mathematical software processes to be taken are as follows:

(i) Make a comparison between the Centroidal Polygon scheme against established schemes Polygon and Harmonic Polygon
(ii) Development of an algorithm
(iii) Details of documentation

The comparison of the CP scheme and previous schemes will be performed using Scilab 6.0 software. Before computing using the software, this study requires to manually compute the solutions in Microsoft Excel, to ensure the CP scheme is computable in Scilab 6.0 software. Both schemes would be compared against the exact solution by tabulating the maximum error between the schemes and the exact solution.

The arrangement of resistors and capacitors in the series circuit equation is shown in Table 1. To ensure usability of the values, all the values and the circuit equation diagrams are obtained from the sources [10–12]. Three different voltage values will be used on each of the circuit equations. The voltages range from low voltages, medium voltages to high voltage values.

Table 1 Set of RL circuit equation problems

RL circuit	Value of Voltage, Resistor, Inductor	Equation
	V = 4 V R1 = 8 Ω L = 4 H	Equation: $\frac{di}{dt} = -2i + 1$ [10] Exact solution: $0.5\left(1 - e^{-t(0.5)}\right)$ $y(0) = 0, 0.1 \leq x \leq 0.5$
	V = 2.4 V R1 = 4 Ω L = 2 H	Equation: $\frac{di}{dt} = -2i + 1.2$ [11] Exact solution: $0.6\left(1 - e^{-t(0.5)}\right)$ $y(0) = 0, 0.1 \leq x \leq 0.5$
	V = 2.4 V R1 = 4 Ω L = 2 H	Equation: $\frac{di}{dt} = -0.5i + 1$ [12] Exact solution: $2\left(1 - e^{-t(2)}\right)$ $y(0) = 0, 0.1 \leq x \leq 0.5$

The source of the equation for the exact solution and equation to be used in the schemes, is the equation that was already validated using Kirchhoff's Law. There are two conditions applicable to Kirchhoff's Law, which are the current law of Kirchhoff, and the voltage law of Kirchhoff. The voltage law of Kirchhoff states that the algebraic sum of instantaneous changes in the potential for the voltage dropping around every closed loop, is equal to zero. Each factor in the circuit equation lowers the voltage [7].

To solve circuit equations, Kirchhoff's Voltage Law demands the voltage drop across each element of the circuit equation. For RL circuit equation, the equation can be shown using Faraday's Law and Lenz's Law, where the voltage dropping E_L across the inductor, is proportional to the instantaneous rate of the change of the current I as in Eq. (5) below:

$$E_L = L\frac{dI}{dt} \tag{5}$$

where the inductance, L, is the constant of proportionality.

3 Results and Discussion

The comparison between results of the three modified Euler schemes with the exact solution is discussed in this topic. Table 1 shows the set of RL circuit equation problems used in this experiment. There are three sets of RL circuit equations with different values for voltage (V), resistor (R), and inductor (I). The solutions in Table 2 were determined from:

$$\text{Error} = [E_x - E_v]$$

where E_x= Exact solution and E_vE_v = Euler's improved value. Exact solution using τ value in a limit time within 0.1 to 0.5. The RL circuit equation is derived to calculate the values of V, R and L using the formula $E_L = L\frac{dI}{dt}$.

Table 2 shows the maximum error between the CP scheme with the P and HP schemes. The results from Table 2 shows that proposed CP scheme gives a more accurate result as compared to P and HP.

CP scheme matched the troubleshooting results for circuit Eq. 1 after observation because it gave an accurate value compared to the other two schemes. In the first scenario with step size, 0.001, the P scheme reported a maximum error of 0.077561, and HP scheme reported a maximum error of 0.087282, compared to the CP scheme which reported a maximum error of 0.070915. It shows that the CP scheme gives a smaller value compared to P and HP schemes. The smaller value reflects a higher accuracy. Although the difference of values between CP and P is substantial, CP still returned higher accuracy.

At step size 0.01, the CP scheme reported a maximum error of 0.072289, compared to the P and HP scheme, which reported maximum errors of 0.078934 and 0.088792 respectively. CP scheme still returned higher accuracy even though the difference in maximum error values between CP scheme and P scheme is small.

At step size 0.1, the CP scheme reported a maximum error of 0.086993 while other schemes reported a maximum error of 0.093450 for P scheme, and 0.104539 for HP scheme. This reflects the better accuracy of the CP scheme even with a higher step size. From the result, it shows that the maximum error of CP scheme is differs slightly from P and HP schemes.

For circuit Eq. 2, the result shows that the CP scheme has given a more accurate value as compared to the other two schemes. The CP scheme produced a smaller maximum error of 0.062616 at a small step size, h = 0.001, as compared to the maximum error of P and HP schemes, at 0.072114 and 0.087097 respectively. It shows that CP scheme returns better accuracy regardless of the step size.

Meanwhile, at large step size, h = 0.1, the CP scheme reported a maximum error of 0.064410 as opposed to P and HP schemes which reported maximum errors of 0.073904 and 0.089012 respectively.

Lastly, at step size h = 0.01, the CP scheme reported a maximum error of 0.083762 while P and HP schemes reported a maximum error of 0.092996 and 0.109232. After testing with all three step sizes, circuit Eq. 2 is suitable to be solved

Table 2 Result for maximum error of RL circuit equation problems

Scheme	Centroidal-Polygon			Polygon			Harmonic-Polygon		
Step size	0.1	0.01	0.001	0.1	0.01	0.001	0.1	0.01	0.001
Circuit Eq. 1	0.086993	0.072289	0.070915	0.093450	0.078934	0.077561	0.104539	0.088792	0.087282
Circuit Eq. 2	0.083762	0.064410	0.062616	0.092996	0.073904	0.072114	0.109232	0.089012	0.087097
Circuit Eq. 3	0.410170	0.387115	0.384940	0.422962	0.400567	0.398437	0.455976	0.434335	0.432237

using CP scheme because it gives better accuracy at all step sizes than P and HP schemes.

The results of the final problem, circuit Eq. 3, is shown in the last row of Table 2. For step size 0.1, the CP scheme was proven to be more accurate even though the difference was small. As shown in Table 2, the maximum error for CP scheme is 0.410170, compared with P scheme and HP scheme, with individual maximum errors of 0.422962 and 0.455976.

When the step size h = 0.01, the CP scheme shows a better result. The scheme returned a maximum error of 0.387115. The maximum error differed slightly from P and HP schemes. As shown in Table 2, P scheme result is 0.400567 and HP scheme is 0.434335.

Lastly, for the small step size h = 0.001, the CP scheme, resulted a maximum error of 0.0384940 compared to 0.398437 and 0.432237 for P and HP schemes respectively. CP scheme is much accurate when compared to HP scheme and also better accuracy against P scheme even though the difference is not substantial.

From these three equations, it shows that the CP scheme is suitable in solving RL circuit equation and returns better accuracy compared to P and HP schemes.

4 Conclusion

The key emphasis of the present analysis is to emphasize the maximum error, which is caused by the RL circuit equations and their variations in speed. The most crucial factor for the electric field is speed, and the accuracy is focused on the Euler's method. The CP scheme is more effective in this research than the P and HP schemes. In this research, the HP scheme is faster than the P and CP schemes in terms of speed and time taken to complete the simulation. It should be mentioned that the CP scheme returns more accurate results as opposed to P scheme and HP scheme, and also faster than P scheme. In conclusion, the CP scheme can serve as a substitute scheme to solve RL circuit equations.

Acknowledgements The authors gratefully acknowledge the University College TATI and author Nooraida Samsudin also would like to thank MOHE for the FRGS grant (FRGS/1/2019/TK04/TATI/03/1).

References

1. Ogunrinde RB (2012) On some numerical methods for solving initial value problems in ordinary differential equations. IOSR J Math 1(3):25-31. https://doi.org/10.9790/5728-0132531
2. Chapra S, Canale R (2006) Numerical methods for engineers, 6th edn., pp 707–726
3. Samsudin N, Yusop N, Fahmy S, Mokhtar A (2018) Cube arithmetic: improving euler method for ordinary differential equation using cube mean. Indonesian J Electr Eng Comput Sci 11 (3):1109–1113. https://doi.org/10.11591/ijeecs.v11.i3

4. Retrieved 2 November 2020, from https://www.ocf.berkeley.edu/~reinholz/ed/08fa_m160/lectures/eulers_method.pdf
5. Mohd Yusop N, Hasan M (2015) Development of new harmonic Euler using nonstandard finite difference techniques for solving stiff problems. Jurnal Teknologi 77(20). https://doi.org/10.11113/jt.v77.6546
6. Yusop N, Hasan M (2017) Comparison new algorithm modified Euler based on harmonic-polygon approach for solving ordinary differential equation. Jurnal Teknologi (Sciences And Engineering) 9(2–11):29–32
7. Nagle R, Saff E, Snider A (2014) Fundamentals of differential equations, 8th edn., pp 117–121
8. Salleh Z (2012) Ordinary differential equations (ODE) using Euler's technique and SCILAB programming. Math Models Methods Modern Sci 4:264–269
9. Alexander C, Sadiku M (2013) Fundamentals of electric circuit equations, 5th edn., pp 259–265.
10. Salford UO (2018) University of Salford. Retrieved August 24, 2018, from http://www.mathscope.salford.ac.uk/tutorials/eleceng/sheet5/sheet-5.pdf
11. Retrieved 2 November 2020, from http://www.fysik.lu.se/fileadmin/fysikportalen/UDIF/Bilder/Lecture7.pdf
12. Tool B, Search P, Designs R, Articles T, Articles I, Equipment T et al (2020) Series resistor-inductor circuit equations | reactance and impedance—inductive | electronics textbook. Retrieved 2 November 2020, from https://www.allaboutcircuit equations.com/textbook/alternating-current/chpt-3/series-resistor-inductor-circuit equations/

Investigating the Relationship Between Medicine Student's Lifestyle and Their Academic Performance

Yahya Almurtadha, Osama S. Mohammed, Hyder Mirghani, and Mukhtar Ghaleb

Abstract The corona virus pandemic encourages the world to look for more medical doctors strengthen educational institutions to understand facilitate their study. In addition to the local citizens, Saudi Arabia hosts millions every year as workers or as religious visitors that increase the demands for qualified medical doctors. Medical school of University of Tabuk is serving Tabuk region by providing medical education and training program. Literature reported that the admission criteria are not enough for predicting the performance of medical students in their future careers and recommended conducting some research studies to address the relationship between lifestyle factors and academic performance of medical students during their medical program. This research aimed to assess the relationship between lifestyle and academic achievements among medical students of University of Tabuk. Two analysis methods were used namely, statistical analysis and data mining prediction analysis. Statistical analysis reported significant relations (P-value < 0.050) between the academic performance and hours spent surfing social media, Gender, Mother level of education, job of the father and Diet lifestyle. Data mining techniques also were used to extract hidden, previously unknown patterns or rules of knowledge to predict the excellent performance. Prediction algorithms reported significant relationship between the excellent performance and slept not more or less than 6–9 h per day, father level of education and having a private car since the campus location is outside the city.

Y. Almurtadha (✉)
Faculty of Computing and Information Technology, University of Tabuk, Tabuk, Saudi Arabia
e-mail: y.murtadha@ut.edu.sa

O. S. Mohammed · H. Mirghani
Faculty of Medicine, University of Tabuk, KSA, Tabuk, Saudi Arabia
e-mail: h.merghani@ut.edu.sa

M. Ghaleb
College of Sciences and Arts—Alnamas, University of Bisha, Bisha, Saudi Arabia
e-mail: mghaleb@ub.edu.sa

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_37

Keywords Lifestyle · Medical students · Academic performance · Educational data mining

1 Introduction

Medical education in Saudi Arabia has shown remarkable improvements. It was started by five colleges and now reaches twenty one medical colleges putting in mind outcome based and community needs [1]. These medical programs gone through two main phases for the purpose of improvement; the first phase with six-year traditional curriculum [1]. The second phase: establishing international partnerships to bond with innovative educational institutes with an intension to acquire quality assurance. There is a strong need to develop qualified doctors to face the new-medical challenges specially knowing that Saudi Arabia hosts millions of people from all around the world every year.

One way to achieve that is to understand the factors that may affect medical students' academic performance to find the possible means to cope or reduce the negative aspects of those factors. Neglecting these factors may lead to medical student's dropout. Pioneers of medical education must address both internal and external challenges affecting medical students with a key challenging factor of the exponential growth in knowledge [2]. These factors are creating an environment full of stress which may affects the learning and cause physical and mental health problems [3]. Recent studies proved that medical students face a higher degree of stress compared with non-medical students of different programs in the university [3]. This has raised the alarm for the compulsion of lifestyle modifications [4]. A successful doctor characteristics include: "feeling happy, satisfied of his achievement and career and managing to live with minimized stress" [5]. Studies during the preliminary year of medical school showed less sleep hours, ignoring physical activity, and neglecting social interactions [6]. These negative modifications of life style showed negative impacts on student's academic performance. Knowing those factors help to understand then design for operative awareness programs or direct for curriculum design. These factors also may help to predict the success of the future doctors. This study aimed to assess the factors disturbing academic performance among medical students of University of Tabuk. The results of this study may help education policy makers to suggest educational strategies to facilitate the prevention of students drop out from medical schools. The structure of this paper will be as followed: Related works and previous studies will be discussed in section II. Section III will give a clear explanation of the methodology including the data collection to form the dataset, the statistical analysis the n finally the data mining analysis to predict the most affective life style features affecting medicine students' academic performance. Last section is the discussion of the results and findings followed by a conclusion followed by the references list.

2 Related Works

Normally, medical schools has their own standards and criteria for admitting new medical students including "academic ability, insight into medicine, keeping updated and do extracurricular activities and interests, personality, self-motivated, and linguistic and communication skills"[7]. A 20 years prospective study in UK found that 47 (9%) doctors were discontinued Medical Register knowing that their grades were lower than those who are still registered [5]. A study at Nottingham University Medical School found that 10–15% per yearly admitted students were encountered to face problems during their medical training [8].

Previous studies reported many factors may affect the medical student's performance during schools. Lievens and his colleagues [9] reported five factors affect medical students performance which are: "Neuroticism, Extraversion, Openness to experience, Agreeableness and Conscientiousness". Rosal et al. [10] warned that emotional misery could lead to be chronic and persistent with dramatic increasing over time. According to Dahlin, Joneborg and Runeson [11] medical students are accompanied with more unhappiness rates than others and women tend to have more stress than men. Literature proved that Women overcome that stress by seeking to get honor degrees to overcome men in the medical fields [7]. "Sex, ethnicity, previous academic ability, personal statements, references, and interviews before selection for medical schools" were reported also as factors associated with success in medical school [7]. Mandal et al. [12] found that "difficulty in understanding medium of instruction, self-assessed depression, sleep disorders and perceived parental and peer pressure and dissatisfaction with career choice" were significantly linked with poor performance.

Other studies suggested some solutions to enhance the academic performance as the study by Stott [13] found that male nursing students may feel alone and barred from academic and clinical perspective so they preferred to engage in the technical aspects of nursing that lead to suggest educational policies to ease the keeping of male nursing students in undergraduate nursing courses. Sambunjak and his colleagues support the concept that mentoring is professed as an important part of academic medicine [14]. Investigating the factors affecting medical students in Saudi Arabia started early. Alfayez, Strand and Carline [15] concluded that "motivations, such as social gains, financial benefits or family wish", were related to lower performance in medical schools. Other researchers later studied several factors and reported them through the last two decades leading to modifications on some programmes. A study by Al-Rubkan et al. [16] supported that the current admission criteria to medical schools in Saudi was sufficient to predict the future performance. This study also suggested "including multiple mini-interviews and the questionnaire for candidate's suitability to follow a problem-based learning curriculum as admission tools". Almost of the previous studies in Saudi Medical schools studied the factors prior to medical enrollment. These factors are essential for screening the candidates to select the expected future students. What is needed

now is to study the in-programme factors that affect the performance of the medical students in Saudi Arabia since several studies recently highlighted that these factors are of great influence especially in a fast growing and modified world.

3 Methods

This research will apply two analysis methods to discover the factors affecting the medical students' performance during their study. The first method is a statistical analysis using STATISTICA software to discover the most-affecting factors among the 23 factors conducted by the questionnaire. The second method is data mining which proves to be efficient techniques to obtain attractive (non-trivial, implicit, previously unknown and potentially useful) patterns or understanding from the enormous amount of data [17]. We used RapidMiner tools for mining the data. This section will start with an overview of the data set used for experiments then description of the two methods used for analysis.

3.1 Data Set

The study was conducted between April and June 2014 at the College of Medicine, Tabuk University. It includes a random sample of medical students in the third (L3), fourth (L4), fifth (L5), and six (L6) academic levels; a self-administered questionnaire was circulated to evaluate demographics, English language entry score, attendance, sitting preference, reading material, study preference, chronic disease, reading style, dietary habit, smoking, weekend educational activity, sleep duration, and average hours spent on social media. Their performance was stratified as "excellent" (CGPA $\geq$ 4/5) or "average" (CGPA < 4/5). Students were asked to complete a self-administered questionnaire. Their Contribution was voluntary.

3.2 Statistical Results and Analysis

We performed statistical analysis using STATISTICA software. The final analysis included 89 students, 57 (64%) of them were males. Their ages ranged from 20 to 25 years with a mean of 22.47 years. (80.9%) of them live with their families, (7.9%) of them live alone, and (11.2%) of them live with their friends. (69.7%) of them used their private car for transportation, (25.8%) of them came with a driver, and (4.5%) of them had shared transportation. (66.3%) of students scored (A) in English language, (23.6%) of them scored (B), and (10.1%) of them scored (C). (88.8%) of the students had good attendance, (9%), (2.2%) of the students had moderate and weak attendance respectively. (69.7%) of the students followed

specified timetable. (37.1%) of the students read from a textbook, (1.4%) of them read from E-book, (25.8%) of them read from lecture note, and (24.7%) of them read from various sources. (62.9%) of the students had excellent performance (GPA $\geq$ 4/5), and (37.1%) of them had "average" performance.

Statistical analysis showed a significant difference (p value < 0.05) between excellent performance and the factors shown in Table 1 with an associated p value for each affecting factor. Time spending on Social Media is a key factor with p value of 0.002934 (data mining algorithms explained in the following section will support this finding). Second in the list is the Student Gender as also proved by similar study [18]. Female students perform better than men due to their intention and extra efforts. Third, Mother education factor gives an indication that mother level of education is affecting the students' performance in the sense that she finds the suitable means to encourage and support him in a positive manners. Similar studies proved the contribution of mothers' education level [19, 20]. This is true since the student should achieve well if they were directed in a proper manner by the parents. Fourth, the job of the father showed association with the academic performance. Finally, Diet lifestyle as proved by similar studies showed an association with the academic performance [21, 22].

4 Data Mining Results and Analysis

Data mining has attracted scientist's attention due to the ability to deal with large datasets that find its way doing many functions. Several researchers applied Data Mining Algorithms to enhance education techniques for various specializations [23]. Tair and El-halees [24] showed the usefulness of using data mining to improve graduate students' performance. Mining the data set requires applying the main processes showed in Fig. 1 namely: data entry, preprocessing, transformation, data mining and results interpretation. We used RapidMiner software for data mining. The following subsections will report the generated rules by data mining algorithms and the findings.

4.1 Random Forest

Random forests are a "grouping of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for

Table 1 Factors with significant contribution to excellent performance

Diet lifestyle	Job of father	Mother education	Gender	Social media	Factor
0.038806	0.02461	0.014157	0.008696	0.002934	p value

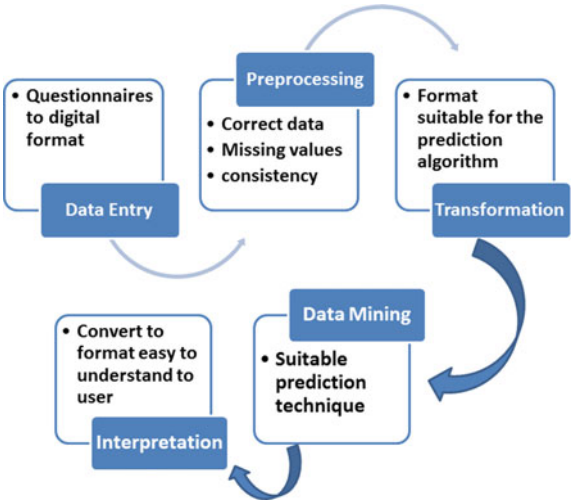


Fig. 1 Data mining processes

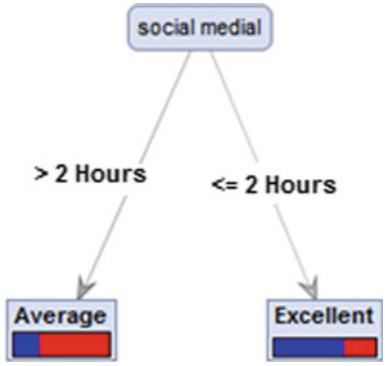


Fig. 2 Random forest rule for excellent performance

all trees in the forest" [25]. Random Forest is a prediction technique under data mining used here which produced two rules as shown in Fig. 2 to predict getting excellent GPA for those students whom use social media. These rules are also illustrated in Fig. 3 associated with the accuracy estimation.

Twenty-five (43.9%) of the excellent group paid ≤ 2 h per day on social media that in accordance with Walsh et al. [26] and Kirschner and Karpinski [27] who concluded that Facebook® users, for example, described having worse GPAs and studying less hours per week than students with no accounts.

```
If social medial > 2 hours/day: Average {Excellent=15, Average=38}
If social medial ≤ 2 hours/day: Excellent {Excellent=23, Average=10}
```

Fig. 3 Random forest rule with accuracy estimation

4.2 Single Rule Induction

Single Rule Induction returns the best prediction rule using exhaustive Breadth First search. Using the medical students' data to predict getting excellent, Single Rule Induction produced this main rule showing that if the student sleeping behavior is between 6 and 9 h per day and his father of education is university level then the student will achieve excellent CGPA. Sleeping less or higher than 6–9 h shows negative affect on the students' performance as proved by similar studies [28, 29].

ConjunctiveRule (prediction model for label CGPA).

(father education = University), (sleep = 6–9) → (CGPA = Excellent).

4.3 Adaboost with Random Forest Regression

Father level of education proved to be a key factor in affecting the son/daughter academic performance as shown in Fig. 4 produced by analyzing this factor independently using Adaboost with random forest regression technique.

father edu > Secondary: Excellent {Excellent = 24, Average = 6}.

father edu $\leq$ Secondary: Average {Excellent = 19, Average = 37}.

4.4 Decision Tree Algorithm

Finally, it is not surprising to observe that medical education has been influenced by indirect factors that directly affect student performance such as the campus location. Decision tree algorithm shows that transportation is an important feature to predict the excellent achievement as illustrated in Fig. 5. This rule states that if the student has a private transportation then he/she has a bigger probability to get excellent CGPA, but if he/she share transportation with others or use the public transportation

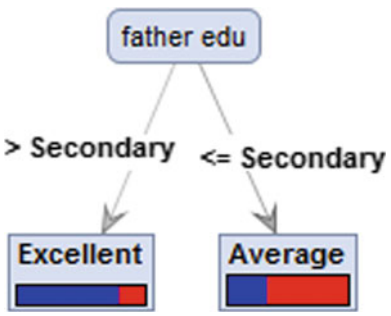


Fig. 4 Father education level rule

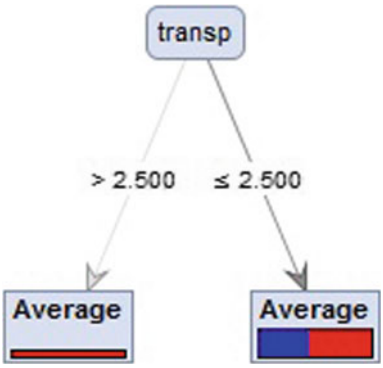


Fig. 5 Transportation as a rule for academic performance

then he/she has a bigger probability of getting average CGPA. This can be explained by the fact that University of Tabuk is a new-established university (since 2006), and the campus is located a little far from the last point of Tabuk city.

5 Conclusion

Several studies tried to investigate the relationship between life style and academic performance of university students. Various admission criteria's are applied to the candidates for medical schools to guarantee the success of selection process of the future doctors. Previous studies reported the drop out of number of students raising the alarm to conduct research to find out the possible reasons. Several factors are affecting medical students' performance including demographic, social, education style, and lifestyle. This research tried to investigate the lifestyle factors that affect medical students of Tabuk University to report the significant relationships associated with excellent achievement.

Two analysis methods applied namely: statistical analysis and data mining prediction analysis. There is statistically significant difference (P-value < 0.050) between the excellent performance and hours spent surfing social media, Gender, Mother level of education, job of the father and Diet lifestyle. Data mining Prediction algorithms reported significant relationship between the excellent performance and slept not more or less than 6–9 h per day, father level of education and having a private car since the campus location is outside the city.

Conducting such studies encourage researchers like [30] to investigate the effect of academic modification to assure the students success in considering the motivation physiological and behavior effects or anxiety and depression as examined by [31]. A future research might be conducted to investigate the performance of the graduated doctors and their contribution to the medical society. Another investigation may consider the effect of pandemic fear such as Corona.

References

1. Telmesani A, Zaini R, Ghazi H (2011) Medical education in Saudi Arabia: a review of recent developments and future challenges. EMHJ 17(8):2011
2. Densen P (2011) Challenges and opportunities facing medical education. Trans Am Clin Climatol Assoc 122:48
3. Yusoff MSB, Rahim AFA, Yaacob MJ (2010) Prevalence and sources of stress among Universiti Sains Malaysia medical students. The Malays J Med Sci MJMS 17(1):30
4. Fiedler PT (2008) Avaliação da qualidade de vida do estudante de medicina e da influência exercida pela formação acadêmica. Universidade de São Paulo
5. McManus I et al (2003) A levels and intelligence as predictors of medical careers in UK doctors: 20 year prospective study. BMJ 327(7407):139–142
6. Tempski P et al (2009) A questionnaire on the quality of life of medical students. Med Educ 43(11):1107–1108
7. Ferguson E, James D, Madeley L (2002) Factors associated with success in medical school: systematic review of the literature. BMJ 324(7343):952–957
8. Yates J, James D (2006) Predicting the "strugglers": a case-control study of students at Nottingham University Medical School. BMJ 332(7548):1009–1013
9. Lievens F et al (2002) Medical students' personality characteristics and academic performance: a five-factor model perspective. Med Educ 36(11):1050–1056
10. Rosal MC et al (1997) A longitudinal study of students' depression at one medical school. Acad Med 72(6):542–546
11. Dahlin M, Joneborg N, Runeson B (2005) Stress and depression among medical students: A cross-sectional study. Med Educ 39(6):594–604
12. Mandal A et al (2012) Factors affecting the performance of undergraduate medical students: a perspective. Indian J Community Med Official Publ Indian Assoc Prev Soc Med 37(2):126
13. Stott A (2007) Exploring factors affecting attrition of male students from an undergraduate nursing course: a qualitative study. Nurse Educ Today 27(4):325–332
14. Sambunjak D, Straus SE, Marušić A (2006) Mentoring in academic medicine: a systematic review. JAMA 296(9):1103–1115
15. Alfayez SF, Strand D, Carline J (1990) Academic, social and cultural factors influencing medical school grade performance. Med Educ 24(3):230–238
16. Al-Rukban MO et al (2010) The ability of the pre-admission criteria to predict performance in a Saudi medical school. Saudi Med J 31(5):560–564
17. Han J, Kamber M, Pei J (2006) Data mining: concepts and techniques (The Morgan Kaufmann Series in Data Management Systems)
18. Masson A et al (2003) Girls are more successful than boys at the university. Gender group differences in models integrating motivational and aggressive components correlated with Test-Anxiety. L'Encephale 30(1):1–15
19. Mushtaq SNK (2012) Factors affecting students' academic performance. Glob J Manag Bus Res 12(9)
20. Raychaudhuri A et al (2010) Factors affecting student's academic performance: a case study in agartala municipal council area. Bangladesh e-J Sociol 7:34–41
21. Florence MD, Asbridge M, Veugelers PJ (2008) Diet quality and academic performance*. J Sch Health 78(4):209–215
22. Al-Kandari F, Vidal VL (2007) Correlation of the health-promoting lifestyle, enrollment level, and academic performance of College of Nursing students in Kuwait. Nurs Health Sci 9(2):112–119
23. Educational Data Mining (2008) In: The 1st international conference on educational data mining. 2008. Montréal, Québec, Canada
24. Tair M, El-Halees AM (2012) Mining educational data to improve students' performance: a case study. Int J Inf Commun Technol Res 2(2)
25. Breiman L (2001) Random forests. Mach Learn 45(1):5–32

26. Walsh JL et al (2013) Female college students' media use and academic outcomes results from a longitudinal cohort study. Emerg Adulthood 2167696813479780
27. Kirschner PA, Karpinski AC (2010) Facebook® and academic performance. Comput Hum Behav 26(6):1237–1245
28. Lemma S et al (2014) Good quality sleep is associated with better academic performance among university students in Ethiopia. Sleep Breathing 18(2):257–263
29. Abdulghani HM et al (2012) Sleep disorder among medical students: relationship to their academic performance. Med Teach 34(s1):S37–S41
30. Raza SA, Qazi W, Yousufi SQ (2020) The influence of psychological, motivational, and behavioral factors on university students' achievements: the mediating effect of academic adjustment. J Appl Res Higher Edu
31. Jenkins PE, Ducker I, Gooding R, James M, Eley E (2020) Anxiety and depression in a sample of UK college students: a study of prevalence, comorbidity, and quality of life. J Am College Health 1–7

Improved Search Pattern with Discrete Wavelet Transform for Video Steganography

Hayfaa Abdulzahra Atee, Abidulkarim K. I. Yasari, and Dalal Abdulmohsin Hammood

Abstract The scope of this research is to overcome the computational complexity of the methods for hiding messages inside video frames using various advanced search procedures. For this purpose, the optimal binary coding scheme with low-complexity frame adjustment has been implemented for wireless video transmission applications. To minimize the computational complexity of concealing the existence of messages and the ability to accurately extract them, the Improved Search Pattern with Discrete Wavelet Transform Based Phase Correlation (ISP-DWT-PC) was computed. The approach has also been used to improve the speed of the hidden data retrieval reliably using a simpler search pattern. This paper emphasizes computer forensics procedures, which acts as a tool for data security, reliability, and authentication in video steganography applications works. The results that have been achieved is surpassed the traditional methods in terms of PSNR, MSE, SSIM, Accuracy, surpassed, sensitivity and specificity.

Keywords Wireless video communication applications · Video steganography · ISP-DWT-PC · Binary coding · Data security

1 Introduction

For industry-standard steganography implementations, several video steganography search protocols have been used recently. The documentation of Advanced Video Coding methodology was issued in 2003 by the International Telecommunication

H. A. Atee
Middle Technical University (MTU), Institute of Administration Al-Rusafa, Baghdad, Iraq

A. K. I. Yasari (✉)
Al-Muthanna University, Engineering College, Samawah, Iraq
e-mail: abidulkarim@mu.edu.iq

D. A. Hammood
Middle Technical University (MTU), Electrical Engineering Technical College, Baghdad, Iraq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_38

Union (ITU-T) and the International Organization for Standardization/International Electro Technological Commission (ISO/IEC) [1]. Advanced search procedures are a safe choice for transforming high video density format to low video density format while transferring or storing data within images. Video coding is a foundational technology used in optical and portable TV, DVD-Video, and remote video-conferencing systems [2]. In steganography applications, the encoder hides text within a video and a decoder retrieves the hidden text from it [3]. To create the compressed bit stream in the graphic files, the video coding technique creates a variety of syntax elements that must be encoded. These syntax elements quantize and compress the transformation coefficients, allowing the decoded data to regenerate the prediction [4, 5]. Details of the compressed data structure [6] and compression tools utilize the duration to encode the full video sequence data. In bit stream decoding, every syntax element and the compressed bit stream within the video is encoded and decoded using advance search procedures [7]. It reverses the coding process and then reconstructs the text inside the video sequences using the inverse transform and quantization coefficients. The extraction of phase correlation-based text has been used to determine the comparative interpreted offsets within the video frames between the correlated data [8]. It is utilized for frame registration and is relying on a frequency domain representation of the data that is measured using Fast Fourier Transforms (FFT). A subset of cross correlation methods that distinguish the phase data from the cross-correlogram Fourier space illustration will benefit from this feature [9]. The accompanying frame illustrates the use of the phase correlation to describe the comparative translated motion between the two Gaussian noise degraded frames, which affects the quality of the data [10, 11]. The main aims of this work are:

- To decrease the computational complexity text hiding by utilizing the optimal binary encoding technique.
- To improve the speed of data retrieval within visual files using the hybrid transformation method with optimum search pattern.

The introduction details the encoder and decoder techniques, while the literature review (Sect. 2) encompass the search techniques, transformation-based techniques, and its implementation and drawbacks. A brief illustration of the conventional search pattern techniques and proposed Improved Search Pattern with Discrete Wavelet Transform based Phase Correlation (ISP-DWT-PC) is discussed in Sect. 3. Sections 4 and 5 concludes the research and recommend possible future research.

2 Literature Review

The topic of masking data inside video frames has been discussed in a lot of literature, and different approaches and methods have been used. The constrained two-bit transform (C2BT) for low complexity motion estimation was used to hide

data. The C2BT reduces the computational complexity of concealed text frames to 2-bit representations. The results of the simulation show that the C2BT algorithm increases the accuracy of the text by 0.35 dB for 2BT-BME and 0.28 dB for C1BT-Block Motion Estimation (BME), respectively [12]. A proposed fast motion estimation with data hiding based on the enhanced forced one-bit transform has been proposed [13].

An upgraded version of C1BT-BME has been reported. The binary transform based BME determines the best Motion Vector (MV) for minimizing computational complexity. It also encourages hardware execution by utilizing the total number of Non-Matching Points (NNMPs) instead of the Sum of Absolute Differences (SAD) [14]. In any case, a Motion Block (MB) size is $N \times N$, and along these lines, the dynamic variety of NNMP ($0 \leq \text{NNMP} \leq N \times N$) diminished to 1/256, which is associated with SAD 8-bit pictures ($0 \leq \text{SAD} \leq 256 \times N \times N$). The higher the NNMP among the present and the previous bit planes, the more reasonable it will be block-most like one another. The data matching error paradigm of the NNMP is lengthy for the purpose of improving the execution of the visual files. The test results demonstrated that the proposed calculation enhanced the retrieval of data by 0.27 dB relative to C1BT-BME, 0.38 relative to 2BT-BME, and 0.67 dB relative to 1BT-BME. While this strategy enhances the data retrieval rate, it also increases the computational complexity. BME procedures dependent on low bit-depth portrayal have been proposed in existing methods to decrease the data retrieval rate.

A Weighted Constraint One Bit Transform Based Motion Estimation (WC1BT-BME) system was proposed to develop a matching execution of the C1BT-BME [15]. The WC1BT-BME methodology is executed using an open-loop devise, and the execution is assessed in terms of the Peak Signal-to-Noise Ratio (PSNR) [13, 16, 17], which helps to retrieve the data efficiently. Instead of the 1BT limitation cover utilized in existing methods, the approach is excellent for a two-bit constraint due to the difference between a pixel and its 1BT threshold [18]. The test results demonstrated that the proposed strategy enhances the text retrieval at a lower error rate of execution of the C1BT-BME, 1BT-BME, MF-1BT, and it outperforms other low bit-depth based text retrieval [19] techniques at the MB level. The Adaptive Block Matching Steganography Algorithm Using Bit-Plane Matching considers the magnitude as well as the directive data of the pixel elements in the MB [20].

A basic versatile BMA was additionally created to alter the search starting point and the search range was extended on the block-to-block premise [21]. The experimental results confirmed that the proposed algorithm reported a lower Mean Square Error (MSE) relative to that of other conventional methods with high complexity. However, in the case of data retrieval, it needs ~ 8 times the number of computations relative to other conventional strategies such as the Three Stage Search (TSS) and the Hybrid Predictive Three Stage Search (HPTSS).

To blindly detect MV-based stego videos, a 54-dimensional feature set utilizing spatio-temporal features of motion vectors is proposed. Two details derive from the theory behind the suggested features. First, because of the use of rate-distortion

optimization methods and belonging to the same rigid object or static context, there are heavy dependencies between neighboring MVs. As a result, MV manipulation will reveal significant details about the discrepancies between each MV and the MVs in neighboring blocks. Secondly, while decoding the Lagrangian multiplier, a lot of MVs in original videos are locally optimal. Unfortunately, the technique suffers from losing some information during the compressing process [22].

In both (PM) and (MV) domains, a universal feature set of video steganography is proposed for the steganalysis technique. The feature set is based on two embedding domains, namely the accuracy of the motion vector (MVC), which can be changed by modifications to either PM or MV [23].

This paper espoused the Improved Search Pattern with Discrete Wavelet Transform based Phase Correlation (ISP-DWT-PC) as a suitable solution for computer forensics, acting as a tool for data security, reliability, and authentication in video steganography applications.

3 Conventional Search Algorithm and Improved Search Pattern with Discrete Wavelet Transform Based Phase Correlation (ISP-DWT-PC)

3.1 Three Stage Search (TSS) Algorithm

The algorithm made the searching regions simple and robust with excellent performance [24]. The text has been embedded in each frame of the block in the visual file and the reference frame discovers the data for the current frame current block and delineates the search window. In the final step, the Sum of Absolute Difference (SAD) value was deliberated for 9 various blocks, which includes 8 edge blocks and one center block. In this step, the minimum SAD block was computed to determine the best match block as per Fig. 1. If the minimum SAD block was found to go to step 3, else it goes to step 2. The following steps describe the TSS:

Step-1:

- Start searching the center
- Set S = 1 around the site (0, 0).
- From 16 sites, analyze the one with lowest cost characteristic.
- If the minimal cost feature appears at the outset, set the motion vector to (0, 0).

Step-2:

- Enter this place to different quest outlets if at one of the eight locations the minimum cost is set at S = 1.
- Check the weights of the adjacent site if it can be used to determine 3 or 5 points.
- Place the motion vector to the next point with the lowest weight.

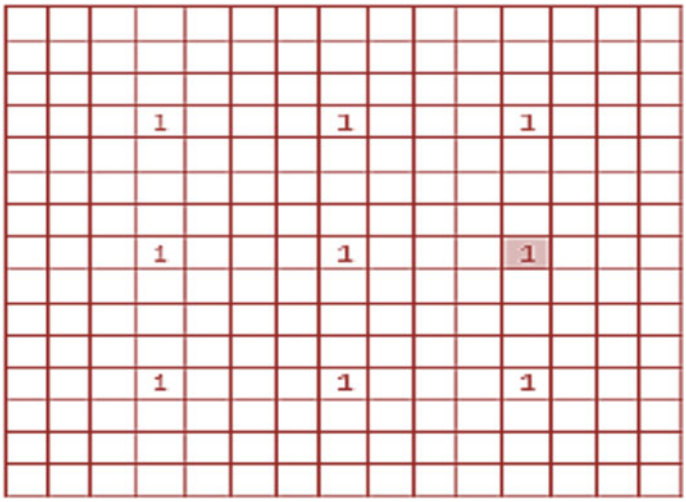


Fig. 1 First stage of three step search

- The normal TSS process takes place if one of the 8 locations in S = 4 is the lowest weight following the first step.

Step-3:

Select from the 9 scanned pages one with a low-cost option.
At the above point, place the current source.
Set the present scale of S to S/2.
Go back l to S = 1 to check the method.
For any macro-block, this algorithm checks 17 points, and in the worst case, 33 positions are to be checked.

Based on these steps, the search is discussed in the following subsections.

3.2 Hybrid Predictive Three Step Search (HPTSS)

This method is based on the Motion Vector (MV) direction the hidden data has been retrieved from. In the neighbor window, the MVs are generally varying deliberately; in most conditions the adjacent block MVs are more or less in the same direction.

The text has been embedded in the blocks of the visual file. If the MVs direction is mostly on the right side, the current block MV is also in the right direction, as shown in Fig. 2. The major awareness of the Hybrid Predictive Three Step (PTSS) search is that it utilizes the left adjacent block MV to confine the search region in

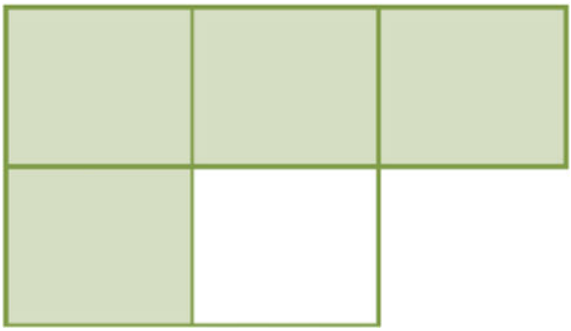


Fig. 2 Neighboring blocks of the current block

the first step of the TSS algorithm. The PTSS utilizes one to four blocks to find the best matching blocks instead of 9 blocks of the TSS algorithm.

3.3 Improved Search Pattern with Discrete Wavelet Transform Based Phase Correlation (ISP-DWT-PC)

In the ISP-DWT-PC, three important methods are implemented, which are designated in the subsequent steps. Improved search can be achieved by SAD measuring the standard function and setting MB as a 16 × 16 macro block where the text is embedded. The hidden data within the video sequence has several properties where when an object moves, the direction of the object and movement speed would not be steady due to its frame rate. General temporal predictions are unsuitable for this work. Instead of the normal MV searching technique, advanced MV with DWT integrated search has been computed with phase correlation to decrease the convergence time. The first MV will solve the issues of text mismatching and lower error rate during the retrieval of hidden message. The creation of the utilization of the indolence speed as well as moving object acceleration for the text is defined as:

$$MVe_j^{(l)} = MVe_{m-1} + (MVe_{m-1} - MVe_{m-2}) \tag{1}$$

where MVe is the motion vector (MV), which can be defined depending on the stability of regular objects:

$$MVe_j^{(2)} = \underset{MV_i}{\operatorname{argmin}} \| MVe_i - \sum_{i=1}^{4} MVe_i \| \tag{2}$$

where MV_i represents the MVs of MB around the current MB, as shown in Fig. 3.

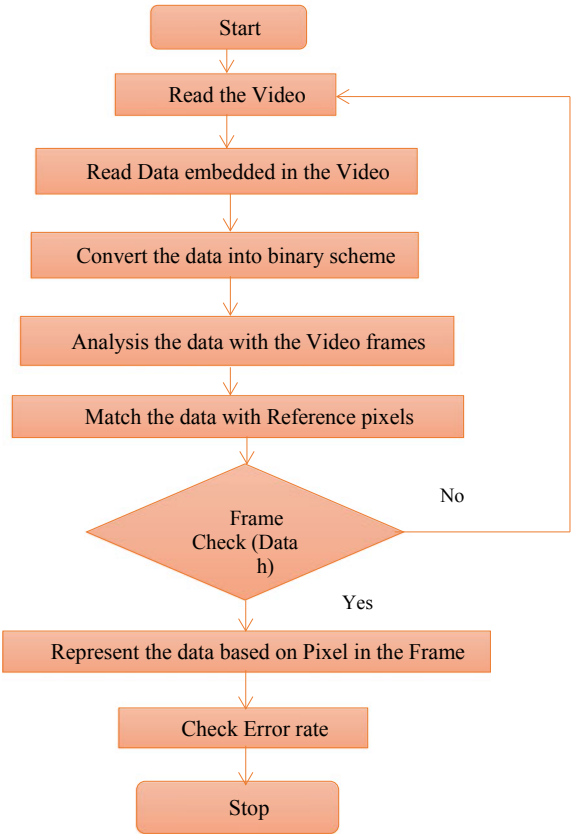


Fig. 3 Flow Diagram of ISP-DWT-PC

The first final MV is selected based on these candidates:

$$MVe_j = \underset{MV_i^{(k)}}{\operatorname{argmin}} \left\{ SAD_{16\times16}(m, n, MVe_i^{(k)}) \right\} \tag{3}$$

where m, n represents the current MB position that is to be coded. The simultaneous increase of the searching region and the convergence speed of the matching is shown in Fig. 4.

Generally, the frame lies on the low frequency band. The degradation to the input raw frame with text embedded and the reference frame to free of high frequency domain is shown in Eqs. (4) and (5).

Fig. 4 The MVs of MB around the current MB

MVe_2	MVe_3	MVe_4
MVe_1	MVe	

$$DD_{\frac{M}{2}\times\frac{N}{2}} = SS \cdot II_{M\times N} \tag{4}$$

$$\widehat{DD}_{\frac{M}{2}\times\frac{N}{2}} = SS \cdot \widehat{II}_{M\times N} \tag{5}$$

where function S_{SA} or subsampling function S_{Sub} with a factor of two horizontally and vertically, respectively. After that, the existing three-step search process is defined with an array step of (4, 2, 1) around

$$\left(\left[\frac{M}{2}\right], \left[\frac{N}{2}\right]\right) + \frac{1}{2}MV_i \tag{6}$$

In

$$\widehat{D}_{\frac{M}{2}\times\frac{N}{2}} \tag{7}$$

This process consumes 25 $SAD_{8\times8}$ to obtain the match point defined in Eq. (8).

$$MVe^{(1)} = F_{TSS}\left(DD_{\frac{M}{2}\times\frac{N}{2}}, D\widehat{D}_{\frac{M}{2}\times\frac{N}{2}}, \frac{1}{2}MV_i\right) \tag{8}$$

This matching point is set back to the reference frame, and search for eight extra points around it to more accurately define the MV in eight $SAD_{16\times16}$, as shown in Eq. (9).

$$MVe^{(2)} = F_{Extra}\left(II_{M\times N}, \widehat{II}_{M\times N}, 2MVe^{(1)}\right) \tag{9}$$

The expressions of $SAD_{8\times8}$ and $SAD_{16\times16}$ are defined as per Eqs. (10) and (11), respectively.

$$SAD_{8\times8}(m,n,MVe) = \sum_{i=0}^{7}\sum_{j=0}^{7}\left|DD\left[\left(\left[\frac{M}{2}\right]+i, \left[\frac{N}{2}\right]+j\right)\right] - \widehat{DD}\left[\left(\left[\frac{M}{2}\right]+i, \left[\frac{N}{2}\right]+j\right)\right] + MVe\right| \tag{10}$$

$$SAD_{16\times16}(m,n,MVe) = \sum_{i=0}^{15}\sum_{j=0}^{15}\left|II[(m+i,n+j)] - \widehat{II}[(m+i,n+j)] + MVe\right| \tag{11}$$

Using addition, subtraction, and shifting right as one operator and $|\cdot|$ as two units allow for the computation of the complexity of the multi-resolution process. If S_{sub} is utilized, the ITSS utilize 15 $SAD_{16\times16}$ only, while TSS utilize 33 $SAD_{16\times16}$. If S_{sub} is selected as the construction operator, the ITSS architecture and TSS

architecture are more or less similar, therefore, the SAD function in [25] steps need to be sampled for $SAD_{16\times16}$.

The existing three-steps search may lead to optimal or complex moving frames of the text due to surface errors. To solve this issue, a random optimization series to the MV is applied, as shown in Eq. (12).

$$MV^{(3)} = F_{RO}(I\left[(m,n),\widehat{I}[(m,n)],MV^{(2)},MV_r\right] \tag{12}$$

where MV_r represents the restricted random MV in the searching region. If MV_r is much better than $MV^{(2)}$, eight additional points need to be checked around MV to determine $MV^{(3)}$. If not, more MVs need to be generated to verify $MV^{(2)}$ until it is at a global optimum. The random optimization is mostly based on a random vector generator. The 2-D gauss random vector generator $G(\mu,\sigma)$ can be used to adjust the mean and variance step wise. It can be defined using the search region D and constants K and Σ. The Pseudo code of the ISP-DWT-PC is described below:

$$k = 0; \mu = MVe^{(2)}; \sigma = \Sigma; MVe^{(3)} = MVe^{(2)};$$
$$While(k < K)$$
$$\{$$
$$MVe_r(k) = G(\mu,\sigma)$$
$$if(MVe_r(k) \notin D)$$
$$continue;$$
$$else\ if\left(SAD_{16\times16}(m,n,MVe_r(k)) < SAD_{16\times16}\left(m,n,MVe^{(3)}\right)\right)$$
$$\{$$
$$MVe^{(3)} = F_{Extra}\left(II_{M\times N},\hat{II}_{M\times N},MVe_r(k)\right);$$
$$K\ +\ +\ ;$$
$$\mu = MVe^{(3)};$$
$$\sigma = \frac{\sigma}{2};$$
$$\}\}$$

A random vector generator using more information in the previous procedures could produce better results.

The discrete cosine transform (DCT) based phase correlation was developed to retrieve hidden data with lower complexity and less MSE. These techniques are defined from the Eqs. (13–16) of the DCT, which are utilized in video coding standards for data embedding, as shown in Eq. (13).

$$X_{cc}(K_1,K_2) = C_1C_2\sum_{m,n=0}^{N-1} x(m,n)\cos\left[\frac{\pi K_1}{N}\left(m+\frac{1}{2}\right)\right]\cos\left[\frac{\pi K_2}{N}\left(n+\frac{1}{2}\right)\right] \quad (13)$$

$$X_{cs}(K_1,K_2) = C_1C_2\sum_{m,n=0}^{N-1} x(m,n)\cos\left[\frac{\pi K_1}{N}\left(m+\frac{1}{2}\right)\right]\sin\left[\frac{\pi K_2}{N}\left(n+\frac{1}{2}\right)\right] \quad (14)$$

$$X_{sc}(K_1,K_2) = C_1C_2\sum_{m,n=0}^{N-1} x(m,n)\sin\left[\frac{\pi K_1}{N}\left(m+\frac{1}{2}\right)\right]\cos\left[\frac{\pi K_2}{N}\left(n+\frac{1}{2}\right)\right] \quad (15)$$

$$X_{ss}(K_1,K_2) = C_1C_2\sum_{m,n=0}^{N-1} x(m,n)\sin\left[\frac{\pi K_1}{N}\left(m+\frac{1}{2}\right)\right]\sin\left[\frac{\pi K_2}{N}\left(n+\frac{1}{2}\right)\right] \quad (16)$$

whereas C_1 and C_2 represented as constants and defined in. Equation (17).

$$C_i = \begin{cases} \sqrt{\frac{2}{N}}, K_i \in \{1,\ldots,N-1\} \\ \sqrt{\frac{1}{N}}, K_i = 0 \end{cases}, i = 1,2 \quad (17)$$

Among the abovementioned DCT transformation, the first transformation (Eqs. 14–16) is utilized in video coding standards. The remaining three transformations can be defined using similar hardware structures as that of the first transform. The complex transformation is defined and implemented using a combination of four real transformations, which are $\{X_{cc}(K_1,K_2) - X_{ss}(K_1,K_2)\} - j\{X_{cs}(K_1,K_2) + X_{sc}(K_1,K_2)\}$, and expressed in Eq. (18).

$$X(K_1,K_2) = C_1C_2\sum_{m,n=0}^{N-1} x(m,n)e^{-j\left[\frac{\pi K_1}{N}\left(m+\frac{1}{2}\right)\right]}e^{-j\left[\frac{\pi K_2}{N}\left(n+\frac{1}{2}\right)\right]} \quad (18)$$

The complex transformation in Eq. (18) is a shift invariant transformation. It can be utilized to execute the phase correlation ME. The $X_p(K_1,K_2), X_c(K_1,K_2)$ denotes the 2-D DCT-based complex transforms signals of $x(m,n)$ and $x(m-m_0,n-n_0)$, respectively. The sizes of both $x(m,n)$ and $x(m-m_0,n-n_0)$ is $N \times N$ and the other shifted version, therefore, $X_c(K_1,K_2)$ is defined in the manner shown in Eq. (19).

$$X_c(K_1,K_2) = X_p(K_1,K_2)e^{-j\left[\frac{\pi K_1}{N}m_0\right]}e^{-j\left[\frac{\pi K_2}{N}n_0\right]} \quad (19)$$

The function of the normalized phase correlation of $X_c(K_1,K_2)$ and $X_p(K_1,K_2)$ is defined in Eq. (20).

$$\frac{X_c(K_1,K_2)X_p^*(K_1,K_2)}{\left\|X_c(K_1,K_2)X_p^*(K_1,K_2)\right\|} = e^{-j\left[\frac{\pi K_1}{N}m_0\right]}e^{-j\left[\frac{\pi K_2}{N}n_0\right]} \tag{20}$$

Equation (21) defines the inverse DCT transform of the normalized phase correlation function.

$$IDCT\left\{e^{-j\left[\frac{\pi K_1}{N}m_0\right]}e^{-j\left[\frac{\pi K_2}{N}n_0\right]}\right\} = \delta\left(m+\frac{m_0}{2}, n+\frac{n_0}{2}\right) \tag{21}$$

The MV data (m_0, n_0) can be defined by locating the impulse peak on the phase correlation plane and the data retrieved based on the above equations with less MSE, Structural Similarity Index (SSIM), and Sensitivity and Specificity.

4 Results and Discussion

The results garnered in this work were compared with the standard matching methods for optimization parameters such as PSNR, MSE, SSIM, Accuracy, and Sensitivity and Specificity using the Mathematical Laboratory Version 14.5.

4.1 Sensitivity and Specificity

Based on the above findings, it is obvious that an important entropy value was obtained using the implemented ISP-DWT-PC process, which boosts the overall recovery efficiency of the secret data inserted into the visible files, as indicated in Eqs. (22 and 23).

$$\text{Sensitivity} = \frac{\text{Number of True Positive(Tp)}}{\sum(\text{True positive(Tp)} + \text{False Negative(Fn)})} \tag{22}$$

$$\text{Specificity} = \frac{\text{Number of True Negative(Tn)}}{\sum(\text{True negative(Tn)} + \text{False positive(Fp)}} \tag{23}$$

According to the above process, the results are shown in Table 1.

In the resulting image, the variance indicates more than the threshold with a pixel size of 255. The intensity difference between the adjacent pixels will be very low in the normal text image. Therefore, 255 will be a much lower pixel size. The network [26] has been continually updated based on the vector of the input and the weight values to help retrieve text data with the help of the image binarization method consisting of 2 values such as black (0) and white (1), with the specific and sensitiveness characteristics shown in Eqs. (22) and (23).

Table 1 Comparative analyses on conventional and proposed methods for sensitivity and specificity

Detection of image	Specificity/Sensitivity			
	Tp	Tn	Fp	Fn
C1BT-BME	1.98	1.41	1.42	1.23
TSS	1.93	1.54	1.32	1.22
HPTSS	1.92	1.52	1.35	1.10
ISP-DWT-PC	1.90	1.37	1.30	1.16

4.2 Accuracy (Acc.)

The accuracy metric was tested using the following equation:

$$\text{Accuracy(Acc.)} = \frac{\text{Number of True positive (Tp)} + \text{Number of True negative(Tn)}}{\sum(\text{Number of True positive (Tp)} + \text{False positive (Fp)} + \text{False negative (Fn)} + \text{True negative (Tn)})} \tag{24}$$

The accuracy's classification obtained from the proposed technical solution is superior to the findings of the previous methods reported in the literature and illustrated in Table 2. The ISP-DWT-PC with binarization analysis increases the maximum probabilistic values.

Table 2 explicitly shows that relative to [27] HPTSS (98.58%), TSS (98.35%), and C1BT-BME (97.22%), the ISP-DWT-PC achieves a higher degree of precision (99.72%). Figure 5 displays the graphical representation of this precision.

4.3 Quantitative Comparison

In addition to the abovementioned comparatives, the PSNR and SSIM metrics were used for objective analysis. The different methods, and the PSNR values in dB and SSIM values are shown in Fig. 6 [28]. This shows that the algorithm suggested re-constructing the secret text in sharp and clear high resolution near the picture of ground truth and provides a 2.77 PSNR and 0.90 increase in structural similarity index (SSIM) relative to those of the current equivalents.

Table 2 Accuracy analysis

Method	Accuracy
C1BT-BME	97.22
TSS	98.35
HPTSS	98.58
ISP-DWT-PC	99.72

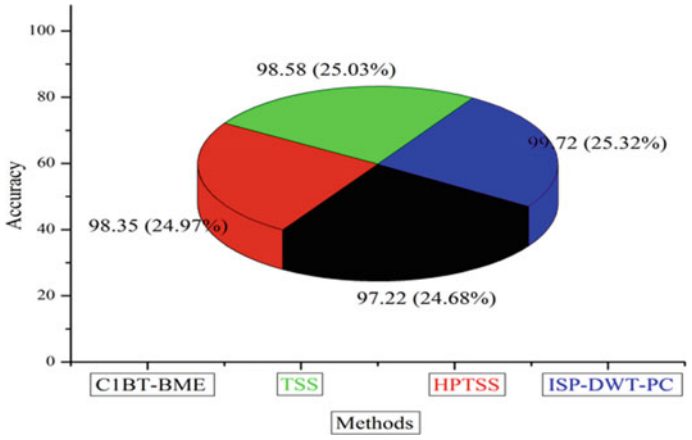


Fig. 5 Accuracy analysis

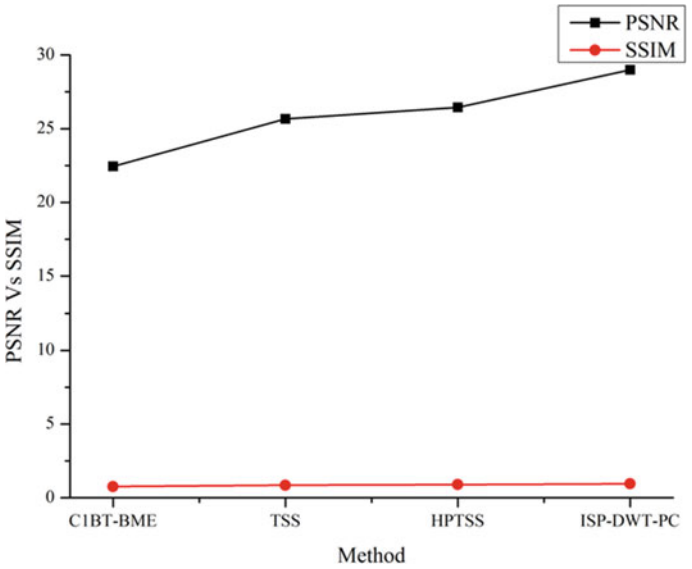


Fig. 6 Methods of CIBI-BME, TSS, HPTSS, and ISP-DWT-PC versus MSE

Hence, from above discussed results, the system effectively minimizes the root mean square error (RMSE) by improving precision, responsiveness, and specificity using state-of-the-art techniques. The RMSE [29] is inversely related to entropy, as the reduction pixel error rate helps maximize the efficiency of the entropy. Figure 7 demonstrates this concept.

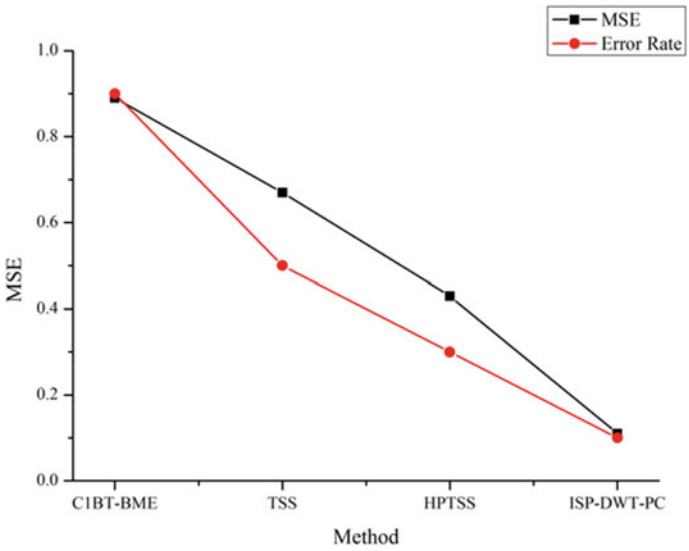


Fig. 7 Methods of CIBI-BME, TSS, HPTSS, and ISP-DWT-PC versus MSE

The MSE marks the collective squared error between the input and output images, as shown in the Eq. (25). The lower the MSE rate, the lower its mistake. Figure 8 shows that the system can efficiently recover data hidden with the most precise PSNR accuracy using the ISP-DWT-PC. In Eq. (26), the $J(m, n)$ and $K(m, n)$ images are the input and output, respectively.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{R^2}{\mathrm{MSE}}\right) \tag{25}$$

$$\mathrm{MSE} = \frac{1}{ab}\left(\sum_{i=0}^{a-1} J(m,n) - \sum_{j=0}^{b-1} K(m,n)\right)^2 \tag{26}$$

The system efficiently identifies the retrieved hidden text in the visual files without changing its characteristics using the ISP-DWT-PC has been validated, and future endeavors in this context is discussed below.

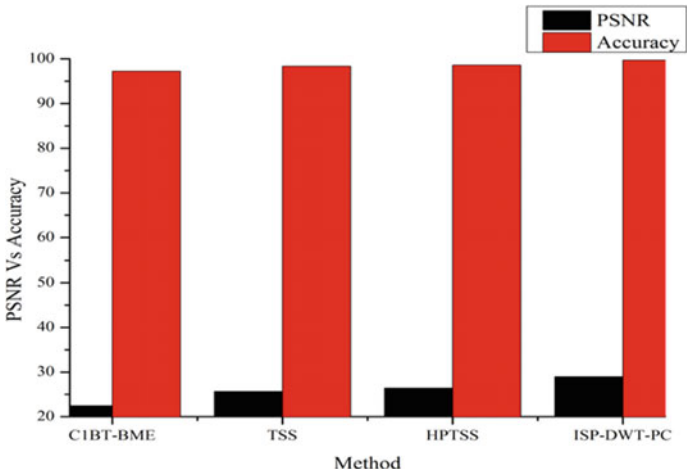


Fig. 8 Classification of PSNR versus accuracy for all methods

5 Conclusion and Future Work

The proposed method of the Improved Search Pattern with Discrete Wavelet Transform based Phase Correlation (ISP-DWT-PC) technique is utilized to reduce the computational complexity of hiding texts. The ISP-DWT-PC improves the speed of the data retrieval inside the visual files using hybrid transformation methods with an optimum search pattern. For future research, the reconfigurable architecture will be developed in a parallel process to increase the accuracy of data retrieval at a lower error rate for computer forensics and video steganography applications. Moreover, the work will be compared with other researchers to ensure the performance of the collected finding.

References

1. Sullivan GJ (2018) Video coding standards progress report: joint video experts team launches the Versatile Video Coding project. SMPTE Motion Imaging J 127(8):94–98
2. Britanak V, Rao K (2018) Audio coding standards, (Proprietary) audio compression algorithms, and broadcasting/speech/data communication codecs: overview of adopted filter banks. In: Cosine-/Sine-Modulated Filter Banks. Springer, pp 13–37
3. Joshi AM (2018) VLSI implementation of video watermarking for secure HEVC coding standard. Cryptographic and Information Security Approaches for Images and Videos
4. Sridhar K, Baskar S, Shakeel PM, Dhulipala VS (2019) Developing brain abnormality recognize system using multi-objective pattern producing neural network. J Ambient Intell Humaniz Comput 10(8):3287–3295
5. Atee HA, Ahmed R, Mohd Noor N, Ilijan AK (2016) Machine learning based key generating for cryptography. J Eng Appl Sci 100(8):1829–1834

6. Fouladi S et al (2017) Encoding, fast and slow: low-latency video processing using thousands of tiny threads. In: 14th {USENIX} symposium on networked systems design and implementation ({NSDI} 17), pp 363–376
7. Dong Y, Jiao W, Long T, He G, Gong C (2018) An extension of phase correlation-based image registration to estimate similarity transform using multiple polar Fourier transform. Remote Sensing 10(11):1719
8. MuhammedShafi P, Selvakumar S, Mohamed Shakeel P (2018) An efficient optimal fuzzy C means (OFCM) algorithm with particle swarm optimization (PSO) to analyze and predict crime data. J Adv Res Dyn Control Syst 10(06): 699–707
9. Liu Y, Liu S, Zhao H, Liu S (2019) A new data hiding method for H. 265/HEVC video streams without intra-frame distortion drift. Multimedia Tools Appl 78(6):6459–6486
10. Boonprong S, Cao C, Chen W, Ni X, Xu M, Acharya BK (2018) The classification of noise-afflicted remotely sensed data using three machine-learning techniques: effect of different levels and types of noise on accuracy. ISPRS Int J Geo Inf 7(7):274
11. Singh P, Raman B (2018) Reversible data hiding based on Shamir's secret sharing for color images over cloud. Inf Sci 422:77–97
12. Duvar R, Urhan O (2018) Adaptive bit-plane selection-based low complexity motion estimation for screen content coding. Turk J Electr Eng Comput Sci 26(5):2206–2217
13. Karampidis K, Kavallieratou E, Papadourakis G (2018) A review of image steganalysis techniques for digital forensics. J inf Secur Appl 40:217–235
14. Chaeikar SS, Zamani M, Manaf ABA, Zeki AM (2018) PSW statistical LSB image steganalysis. Multimedia Tools Appl 77(1):805–835
15. Hussain M, Wahab AWA, Idris YIB, Ho AT, Jung K-H (2018) Image steganography in spatial domain: a survey. Signal Process Image Commun 65:46–66
16. Atee HA, Yasari AKI, Ahmad R, Noor NM (2019) Text in image steganography based on a dynamic non-sequential least significant bit technique in grayscale and rgb images
17. Atee HA (2018) An improved chaotic radial basis resonance theoretic neural network integrated with genetic algorithm for enhancing security in image transmission. J Eng Appl Sci 13(9):3266–3270
18. Kadhim IJ, Premaratne P, Vial PJ, Halloran B (2019) Comprehensive survey of image steganography: techniques, evaluations, and trends in future research. Neurocomputing 335:299–326
19. Shakeel PM, Baskar S, Dhulipala VS, Jaber MM (2018) Cloud based framework for diagnosis of diabetes mellitus using K-means clustering. Health Inf Sci Syst 6(1):16
20. Long M, Li F (2018) A formula adaptive pixel pair matching steganography algorithm. Adv Multimedia 2018
21. Kim J, Park H, Park J-I (2017) Image steganography based on block matching in DWT domain. In: 2017 IEEE international symposium on broadband multimedia systems and broadcasting (BMSB), IEEE, pp 1–4
22. Ghamsarian N, Schoeffmann K, Khademi M (2020) Blind MV-based video steganalysis based on joint inter-frame and intra-frame statistics. Multimedia Tools Appl. https://doi.org/10.1007/s11042-020-10001-9
23. Zhai L, Wang L, Ren Y (2020) Universal detection of video steganography in multiple domains based on the consistency of motion vectors. Trans Inf Forensics Secur 15: 1762–1777. https://doi.org/10.1109/TIFS.2019.2949428
24. Biswas B, Mukherjee R, Chakrabarti I, Dutta PK, Ray AK (2018) A high-speed VLSI architecture for motion estimation using modified adaptive rood pattern search algorithm. Circuits Syst Signal Process 37(10):4548–4567
25. Petersen S, Carlsen S (2011) WirelessHART versus ISA100.11a: the format war hits the factory floor. IEEE Ind Electron Mag 5:23–34. https://doi.org/10.1109/MIE.2011.943023
26. Shakeel PM, Baskar S, Dhulipala VS, Mishra S, Jaber MM (2018) Maintaining security and privacy in health care system using learning based deep-Q-networks. J Med Syst 42(10):186

27. Gordon S, Chin D (2017) Method and system for generating a transform size syntax element for video decoding. US
28. Hemalatha J, Devi MK, Geetha S (2019) Improving image steganalyser performance through curvelet transform denoising. Clust Comput 22(5):11821–11839
29. Jridi M, Napoléon T, Alfalou A (2018) One lens optical correlation: application to face recognition. Appl Opt 57(9):2087–2095

Lightweight Cryptographic and Artificial Intelligence Models for Anti-smishing

Heider A. M. Wahsheh and Mohammed S. Al-Zahrani

Abstract The utilization of Short Message Service (SMS) in smartphones has expanded quickly during recent years. As text informing is the most prudent choice accessible to arrive at masses, it is broadly utilized for offering buy focal points and as a promoting medium. Roughly billions of instant SMSs are sent each hour worldwide, so the expanding utilization of these messages makes it profoundly suitable for assailants. Over 90% of SMS messages are opened within a few seconds, which shows SMS's importance, making it the focus of cyber-attacks. Smishing is one of the cybercrime types that combine SMS and phishing in which trespassers send SMS containing suspicious links and content to the victims. Usually, smishing directs users to harmful websites to silently download malware or show interfaces like legitimate sites to attract users to fill in their sensitive information, i.e., passwords. This study proposes two automated models for detecting smishing by employing Lightweight cryptographic and artificial intelligence (AI) techniques. For the first model, and since there is no official validation stage in SMS construction, we developed our first contribution using a lightweight cryptographic SMS model to protect SMS and detect possible phishing or fake content. We select the standard, popular signature schemes, and evaluating the performance according to time and signature size overhead. The results showed the proposed model's effectiveness to guarantee the main security goals, authentication, integrity, and non-repudiation. Our second model deals with legacy SMS, which doesn't follow the validation stage, so we collected the SMS dataset for an AI model and manually labeled their classes as phishing or legitimate according to their link content. We preprocessed the SMS dataset to extract the link-based distinguishing features, and we applied feature selection techniques to obtain the best performance. We experimented with two artificial intelligence classifiers and

H. A. M. Wahsheh (✉)
Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al-Ahsa 31982, Saudi Arabia
e-mail: hwahsheh@kfu.edu.sa

M. S. Al-Zahrani
Department of Computer Networks and Communications, College of Computer Science and Information Technology, King Faisal University, Al-Ahsa 31982, Saudi Arabia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_39

evaluated their detection accuracy results using accuracy, precision, recall, and F-score metrics. Random forest classifier yielded the best detection results with an accuracy of 97%.

Keywords SMS · Smishing · Lightweight cryptographic · Link-based · Phishing · Artificial Intelligence

1 Introduction

Phishing links are the forged web pages that are made by attackers to imitate legitimate web pages. The majority of these web pages have high visual similitudes to trick their victims. Phishing is an illegal technique that intends to deceive users and steal their individual and sensitive data. The effect of phishing is large due it includes the danger of data fraud and financial risks. Phishing attacks become a major cybersecurity issue for web-based banking, payments, social media, and e-commerce activities [1].

Although SSL and HTTPS techniques are developed to provide safe interaction between users and servers, there are some ways to violate this objective [2]. Let's Encrypt is offering HTTPS certificates without a real check of the web page content. As a result, in 2019, more than 74% of phishing web pages were employing HTTPS compared to 32% of this usage by two years ago [3].

Cyber-attacks are utilizing the public attention of the COVID-19 crisis that was raised in late 2019. It has been recorded that phishing attacks expanded by 667% in one month (between February and March 2020). New phishing web pages were designed to be similar to the World Health Organization (WHO) website. Attackers ask victims to click on claimed medical reports links that direct them to malicious pages or attract them to click on the financial donations links supporting efforts against COVID-19 pandemic to steal their credit card details [3, 4].

Recent statistical reports regarding text message marketing present that 65% of the world population can communicate by Short Message Service (SMS). Researchers believe that this percentage will increase to reach 78% of the world population by the end of 2025. Mobile SMS marketing is read on average by less than 5 s by 45% response rate compared with email offers, which is at just 6%. Along with individual characteristics, SMS messages are more comfortable to receive and read, comparing to emails. SMS marketing stats present that users rapidly respond to a text while spending an hour and a half to respond to an email with the same content. SMS features highlight the significant role of SMS usages among various users' daily applications. For example, 76% of users prefer to employ SMS as alerts for their necessary appointments [5].

The low cost of SMS services by telecom organizations has prompted the expanded SMS [6]. An illegal (spam) message is commonly any unwanted message shipped off the client's mobile device. Spam messages incorporate ads, free services, advancements, grants, and so on. Individuals utilize SMS messages to

convey as opposed to SMS that there is no need for web association and is essential and productive while sending an SMS message. SMS is probably the least expensive approach and can be considered the easiest method to perform phishing attacks. The victim's mobile device contains delicate and individual data like private financial details and login information. Attackers find various approaches to obtain this data from victims' mobile devices by utilizing the particular medium to contain malicious links. Then we will have a new phishing technique corresponding to the used medium. For example, Malicious QR code links called QRishing [7] and Malicious SMS links called smishing which is perhaps the quickest and straightforward ways to spread phishing links with users [8].

Smishing is SMS with phishing content, which is more mainstream these days wherein attacker sends phishing URLs through SMS and requests victims to click on URLs to deceive them and try to obtain their private and sensitive information [6, 9]. Figure 1 presents an example of smishing during the COVID-19 crisis that asks the client to recharge his balance "easily and securely" by visiting the phishing web page.

This study developed a lightweight cryptographic SMS model to protect SMS and detect possible phishing or fake content. We chose the standard, popular signature schemes and evaluating the performance according to time and signature size over-head. Besides, this study collected an SMS dataset and manually labeled their classes as phishing or legitimate according to their link-based features and user feedback. We preprocessed the SMS dataset to extract the link-based distinguishing features, and we applied feature selection techniques. We experimented with two artificial intelligence classifiers and evaluated their detection results.

The remaining part of this paper is arranged as follows: Sect. 2 presents a summary of the smishing detection related work. Section 3 illustrates our adopted framework. Section 4 explains the conducted experiments and compares the results. Section 5 shows the conclusion.

2 Related Work

Smishing is a particular case of phishing attacks, in which attackers employ SMS as a medium to carry the phishing URLs and deceive victims into visiting these malicious pages. This section reviews some existing anti-smishing techniques to understand the current anti-smishing trends.

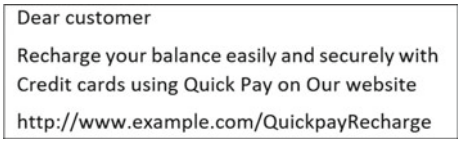


Fig. 1 Smishing quick pay recharge example

The study of [10] proposed an unprecedented method of phishing detection over SMS. It utilized the oversampling approach and the technology of feature optimization to enhance SMS phishing detection accuracy. "Binary Particle Swarm Optimization (BPSO)" algorithm is applied to examine and adopt the best combination of three types of features, including Linguistic Inquiry and Word Count, topic features, and token features. The results indicated that feature optimization and the oversampling method showed a promising performance and enhanced SMS phishing detection accuracy. The optimal accuracy of the suggested approach was 99.01% for an average of features of 86.6. In [11], the authors focused on filtering SMS spam by utilizing an automated framework based on various classifiers. They categorized the classifiers into three sets due to their working, for instance, decision tree-based, ensemble-based, and others. They then assessed the classifiers using some metrics like precision, accuracy, recall, AUC-ROC, and f-score. The method that achieved the optimal performance was random forest along with the chi-square feature selection.

The study of [12] tried to overcome SMSSec protocol limitations by introducing a new SMSSec protocol and then implemented the android platform mobile app depending on that suggested protocol. Throughout the implementation process, the simple protocol also other parts like database access are preserved separately to enable the reusing of the protocol in other apps. Comprehensive analysis of security has demonstrated that the use of mobile apps for sending SMS guarantees mutual authentication, data integrity, non-repudiation, and confidentiality.

The study of [13] utilized a machine learning-based method for spam detection seeking more accurate detection processes. They used various classifiers of machine learning, for instance, K-nearest neighbor (K-NN), decision tree (DT), and Logistic regression (LR). These classifiers are employed for the classification of spam and legitimate messages in the communication of mobile devices. They also used the + e dataset for testing and examining the method. The results indicated a higher LR performance in comparison with DT and K-NN. Additionally, the results showed that LR also attained the highest level of accuracy of 99%.

In [14], a model named' Smishing Detector' is proposed to recognize smishing messages and curtail the false-positive results in each probable step. The suggested model comprised four modules: Source Code Analyzer, SMS Content Analyzer, Apk Download Detector, and URL Filter. Besides, a prototype of the suggested system had been developed and validated by conducting experiments on some SMS datasets. The experiment's results indicated a total of 96.29% of accuracy.

The study of [15] presented a classification of the techniques of anti-phishing on mobile devices and the mitigation techniques available to be used on mobile devices. It aimed to propose an informative model to recognize contemporary solutions for decreasing mobile phishing attacks and demonstrated their future challenges and limitations. The results indicated that each mitigation approach has a particular scope and limitation. The study of [16] suggested a hybrid deep-learning method for SMS spam detecting reasons. The model which had been suggested is based on an integration of both CNN and LSTM deep-learning methods. The model aimed to deal with SMS, which is written in a mix of Arabic or English. The

experiment results showed that the model of both CNN-LSTM achieved an accuracy of 98.37%, which is considered a good rate.

In [17], authors adopted a novel method of machine-learning for SMS spam detection. The method involved two significant stages, the feature extraction, and the decision-making. F-measure and accuracy tested the proposed method, where the result achieved elevated detection rates in both classification accuracy and F-measure.

The study of [18] proposed a novel approach to filter spam from smishing messages. The suggested system contained two stages; the first involved filtering the spam and non-spam SMSs, while the second involved filtering smishing from spam messages. The performance evaluated on machine learning classifiers with a dataset containing spam and non-spam SMSs. The results showed that the utilized approach gained an accuracy of 94.9% in detecting spam messages and an accuracy of 96% in filtering smishing messages.

The study of [19] proposed an anti-phishing model based on the correlation algorithm; called 'SmiDCA' (SMIshing Detection). The model was used to collect varied smishing messages from different sources.

In [20], the authors presented an overall review of the phishing attacks and anti-phishing solutions. They stated that the phishing attacks are divided into two groups; malware-based and social engineering phishing attacks. Simultaneously, the anti-phishing solutions are distinguished into two different kinds; phishing detection and prevention. They conducted a comparison between all kinds of phishing detection solutions and found that academic phishing detection schemes are the most useful ones.

The study of [21] suggested a data-driven model based on deep-learning (multilayer perception) to detect phishing webpages. The used dataset that contained information on 10,000 webpages and ten attributes was generated from "Kaggle". The suggested model had performed at 93% of test accuracy and almost 95% of training accuracy.

In [22], the authors provided an inclusive analysis of the exploitation and solutions of phishing attacks on mobiles. They argued that no single technique is sufficient alone for phishing detection. Moreover, they stated that phishing detection in the mobile environment is considered an outright challenge for more research and development.

3 Smishing Detection Framework

Several phishing detection methods and approaches share general procedures. Still, they differentiated in significant sub-procedures, such as adopting the latest detection features and being independent and not relying on any external source that may lead to detection delay. In this study, our framework consists of the following two main phases:

- Lightweight Cryptographic Model: since there is no official validation stage in SMS construction, we select standard digital signature schemes and evaluate the performance. We can summarize the security requirements in three main points as follows:
  - o Authentication: authenticate the SMS sender.
  - o Data Integrity: Make sure that data has not been modified.
  - o Non-repudiation: Make sure that the SMS sender cannot deny the authorship of the SMS it has created.
- Artificial Intelligence Model: this model consists of the following main sub-points:
  - o Data collection we have collected 30,000 links divided equally into 15,000 phishing links and 15,000 legitimate ones from various sources.
  - o Link-based features extraction and selection: we developed a tool to extract the link-based proprieties according to the recently recommended features and applied a feature selection technique to employ the most optimal features.
  - o Machine learning classification and comparisons: we used and compared two ensemble learning classifiers, bagging, and random forest, to find the best technique to detect smishing.

3.1 Lightweight Cryptographic SMSs

In this context, we first present a summary of cryptography and information security terms and algorithms. Then we will discuss the proposed SMS protection model.

Public key cryptography (asymmetric) is a cryptographic system with two keys; public and private. This method is widely used in data encryption and authentication [23]. At the same time, Hash Function takes a message as input and returns a fixed-size value called "hash." A hash function is a one-way operation; it is tough to get the original data by processing the hash value. It is also impossible to create two messages with the same hash value by using secure and robust hash functions.

Digital Signature is a security scheme that employs public-key cryptography to validate authentication, integrity, and non-repudiation of a message, it computes the hash and then signs it by the private key [23].

Improving the SMSs protection with digital signature prevents phishing and fake SMS content. In an open environment, institutions and companies must configure themselves using Public Key Infrastructure (PKI). They must ask to generate their key pairs, public and private keys, from the certificate authority (CA), the trusted agency responsible for distributing digital certificates. Figure 2 presents digital signature steps to send a message.

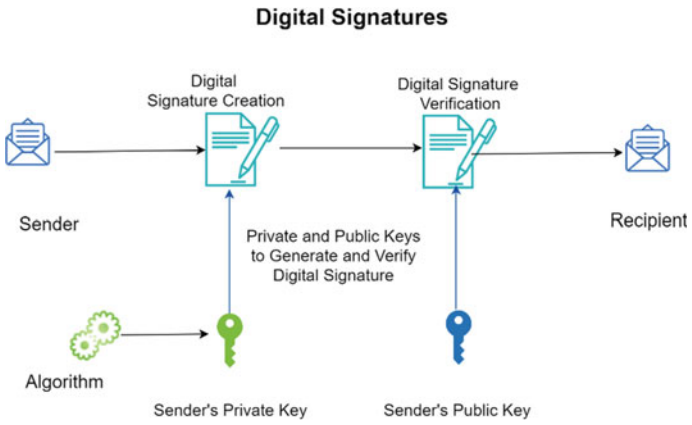


Fig. 2 Digital signature main steps

We assume that if the client wants to open a bank account, this bank will configure the bank public key inside the client's smartphone using our model. According to Fig. 2, the bank will digitally sign the SMS by computing the hash value of the content and sign it using the private bank key. Our model will verify the signature using the stored bank public key when clients receive the SMS and get the hash value. The model will then compute the hash value for the received SMS and compare two hashes that must be the same. In this way, our model guarantees the sender authentication.

Suppose some of the attackers intercept the SMS between the bank and the customer and modify its contents. In that case, the hash value of the modified content will differ from the hash value signed using the private key. So our model guarantees data integrity. Therefore the verification process will reveal that the signature is not identical to the SMS content.

In Fig. 2, if the phishers want to send the phishing SMS and claim that they are the bank, even if they use their digital signature, the result of the client-side verification process will fail because the phishers do not have the bank's private key. Besides, putting their signature, thus exposing them to legal liability. Moreover, if phishers take any digital bank signature from the previous SMSs, the client-side verification process will fail because each message has a unique hash value.

In some cases, phishers embed legitimate links and ask users to download their apps from the Google play store or App store. These applications are not malicious, but they obtain all permissions and penetrate users' privacy when they are run on users' devices. Phishers can use these apps later to get passwords or sensitive data from users when they are run on users' devices [24, 25]. So our model prevents this case because of the non-repudiation and legal responsibility.

3.2 Data Collection for AI Model

A balanced dataset includes an equal or almost equal number of instances from the legitimate and phishing class label. If the instances from one of the classes out-number the other (imbalance), the data is skewed in favor of one class. This may cause training problems because the majority class can control the minority class, and the classifier may only learn one concept alternatively of two different concepts. So, this study adopted a balanced dataset, and we have collected 15,000 phishing links mainly from [26, 27] and the published phishing datasets [28]. At the same time, we have collected 15,000 legitimate URLs from the white-list collections by [28–31].

3.3 Features Extraction and Selection

In this study, we adopt analyzing the link-based features to avoid the network delay because of visiting the URL and collect the content page features or asking for features from external sources such as WHOIS servers. Phishing attackers are employing the shorting URL services such as [32, 33] to attract users to click and visit their links. We explore the recent phishing link-based features [34] and develop a specific tool to extract link-based features after getting the full expanded URLs. Using a single element is not appropriate for the prediction process, and using all features is not preferred too.

We use the feature selection method to find the best detecting best features and exclude the worst ones. Feature selection is considered an important issue. It illustrates which features are relevant for the prediction purpose (i.e., phishing or legitimate) [35].

The feature selection filter method is focused on providing ranks for the features according to their priorities. Features with high priority are ranked first to identify the relevant features that can be used for distinguishing different classes [35, 36]. We applied the correlation attribute evaluator with a Pearson correlation coefficient to find the correlation values between –1 and 1 as negative and positive to the target class (i.e., phishing or legitimate). This method's advantages are that it does not require high computation and by-passes the over-fitting, and is proven to do well for particular data sets. Table 1 presents the adopted 22 features, and their description form [34].

Table 1 The adopted link-based features [34]

ID	Attribute	Description	Type
1	Length of Hostname	The total characters in the hostname	Numeric
2	Queries	The number of queries in the URL	Numeric
3	Dashes1	The number of dashes in the hostname	Numeric
4	Dashes2	The number of dashes in the URL	Numeric
5	Dots	The number of dots in the URL	Numeric
6	And &	The number of (&) in URL	Numeric
7	Hash	The number of hash (#) in URL	Numeric
8	Percentage %	The number of (%) in URL	Numeric
9	Underscore	The number of (underscore) in the URL	Numeric
10	Tilde	The number of (∼) in URL	Numeric
11	@	The number of (@) in the URL	Numeric
12	PathLength	The number of characters in path of URL	Numeric
13	Attractive words	The number of attractive words in URL*	Numeric
14	NumNumericChars	The number of numeric characters in the webpage URL	Numeric
15	Length of URL	The total characters in the URL	Numeric
16	QueryLength	The number of characters in query part of URL	Numeric
17	Domain Inside Sub	Checks if Top Level Domain (TLD) is a part of a subdomain	Binary
18	IP Address	Checks if IP address in the hostname of URL	Binary
19	HttpsInHostname	Checks if HTTPS is obfuscated in hostname part of URL	Binary
20	NoHttps	Checks if HTTPS exist in URL	Binary
21	DoubleSlashInPath	Checks if "//" exist in the path of URL	Binary
22	Domain-Paths	Checks if TLD is in the path of URL	Binary

*For example: (secure), (account), (login), (signin), (banking), (confirm)

4 Experiments and Evaluation Results

This section is divided into two parts. The first is dedicated to assessing the size and time overhead for the lightweight cryptography model, and the second part is devoted to evaluating the artificial intelligence model.

4.1 Cryptographic Model Structure, Time and Space Signature Overhead

We develop a lightweight cryptography model using the Elliptic Curve Digital Signature Algorithm (ECDSA), a public-key cryptographic algorithm commonly used for digital signatures. It is a variant of the Digital Signature Algorithm (DSA),

based on elliptic curves' structure [23]. We followed the recent lightweight cryptographic instruction and used ECDSA with a key length of 128 bits and SHA-256 as a hash function [37].

Since encoding data in SMS use Base64, we define a reserved keywords structure to define the type of encoded data, and the control data follows:

- Creator ID (ID): the unique value that identifies the SMS creator/sender.
- Digital signature (DS): the digital signature on the SMS content.
- Data: contains SMS content.

Since one SMS can carry data up to 160 characters, so to increase the size, we utilize in our model the usage of three SMSs, to avoid converting it to Multimedia Messaging Service (MMS). We dedicate one SMS to hold the security control data that encode ECDSA digital signature with a key length of 128 bits, which needs 76 characters. The remaining control data characters will cover the reserved keywords structure. Creator will have two SMS to encode the real SMS content. Time overhead for ECDSA signature verification requires less than 30 ms using modern smartphones as it has been computed in recent studies [38, 39].

4.2 *Artificial Intelligence Models Evaluation*

As we mentioned in the related work section, machine learning (ML) techniques are widely utilized in phishing detection. Ensemble learning is an ML technique focused on building robust classifiers from weak classifiers and using the robust classifier to classify the instances. Ensemble methods couple many classifiers to form a better classifier with high performance. Two types of ensemble classification method, bagging, and random forest [36]. We used 66% of the data as a training dataset and the remaining 34% as a testing dataset. We compared the yielded results of bagging and random forest (RF) through the prediction quality measurements; True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN), Precision (P), Recall (R) and F-Measure (F-M), as shown in formulas 1–4 [40].

$$Accuracy_i = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$Recall_i = \frac{TP}{TP + FN} \tag{2}$$

$$Precision_i = \frac{TP}{TP + FP} \tag{3}$$

$$F - measure = \frac{(2 \times TP)}{(2 \times TP) + FP + FN} \tag{4}$$

Bagging (bootstrap aggregation) is an ensemble machine learning that is useful for regression and classification. It utilizes many weak models and then aggregating the predictions and chooses the best forecast [41]. Bagging classifier yielded high accuracy results of 92.9% with an error rate of 7%, Table 2 explores the detailed results.

As shown in Table 2, the bagging classifier yielded high prediction results for both legitimate and phishing URLs. A confusion matrix is a particular table design that provides visualization of an algorithm's performance, typically in supervised learning classifiers. Every row presents the actual instances class label in the confusion matrix, while every column shows the instances' predicted class label.

The bagging classifier correctly classified 9477 instances and failed to classify 723 samples to their correct class label, as shown in Table 3.

Random forest (random decision forests) is a popular, flexible, easy to use an ensemble learning algorithm that belongs to the supervised learning technique and uses for classification and regression. It consists of several decision trees on various subsets and computes their accuracy to enhance the final predictive accuracy and prevent the overfitting problem [42]. Random forest enhanced the accuracy results of bagging classifier and obtained 97.2% with an error rate of 2.8%, where Table 4 shows the detailed results.

Table 2 Detailed accuracy results for bagging classifier

Class	TP rate	FP rate	Precision	Recall	F-M
Legitimate	0.914	0.056	0.943	0.914	0.928
Phishing	0.944	0.086	0.916	0.944	0.930
Weighted Avg.	0.929	0.071	0.930	0.929	0.929

Table 3 Confusion matrix of bagging classifier

Actual class	Predicted class	
	Legitimate	Phishing
Legitimate	4691	440
Phishing	283	4786

Table 4 Detailed accuracy results for the random forest classifier

Class	TP rate	FP rate	Precision	Recall	F-M
Legitimate	0.971	0.027	0.973	0.971	0.972
Phishing	0.973	0.029	0.970	0.973	0.972
Weighted Avg.	0.972	0.028	0.972	0.972	0.972

Table 5 Confusion matrix of random forest classifier

Actual class	Predicted class	
	Legitimate	Phishing
Legitimate	4980	151
Phishing	137	4932

Table 5 presents the confusion matrix of random forest, where it successfully predicts 9912 instances to their correct belonging class label.

Random forest enhances bagging accuracy results because it enhances the trees by dividing random subset of features. This indicates that the model considers only a small subset of features at each split of the tree rather than all of the model's features. From the set of available features x, a subset of y features (y = square root of x) is picked randomly. This is vital so that variance can be averaged away. So, if the data set includes a few powerful predictors. These predictors will consistently be chosen at the trees' top level to have very related structured trees to be highly correlated.

5 Conclusion

Recently, attackers have increased in designing smishing (SMS includes phishing URLs) to appear legitimate and identical to the official messages users usually receive from different parties such as the telecommunications company, bank, or medical institution. These messages include a URL, which asks the user to fill in data to exploit users. When users click and visit these links, they will find a website that may appear real, but it is a phishing web page since attackers aim to collect user data from exploiting it. This study proposed two automated models for detecting smishing; where the first we employ a lightweight cryptographic SMS model to protect SMS and detect possible phishing or harmful content. We used standard signature schemes and evaluated the security size and time overhead. The proposed model's results produce effectiveness to guarantee authentication, integrity, and non-repudiation security requirements. Then we collected a dataset of phishing and legitimate URLs. We extracted the link-based features and applied the feature selection method. The conducted experiments used two ensemble learning classifiers; bagging and random forest. The results showed that random forest yielded the best results, and it is the recommended model for anti-smishing.

Acknowledgements The authors acknowledge King Faisal University for the financial support.

References

1. Prasad R, Rohokale V (2020) Cyber security: the lifeline of information and communication technology. Springer
2. Ukrop M, Kraus L, Matyas V, Wahsheh HAM (2019) Will you trust this tls certificate? Perceptions of people working in it. In: Proceedings of the 35th annual computer security applications conference, pp 718–731. ACM, USA
3. European Union Agency for CyberSecurity (ENISA). https://www.enisa.europa.eu, last accessed 2021/01/10
4. World Health Organization (WHO). https://www.who.int/about/communications/cyber-security, last accessed 2021/02/12
5. SMS Marketing Stats. https://99firms.com/blog/sms-marketing-stats/#gref, last accessed 2021/01/12
6. Choudhary N, Jain AK (2017) Towards filtering of SMS spam messages using machine learning based technique. In: Proceedings of the international conference on advanced informatics for computing research, pp 18–30. Springer, Singapore
7. Focardi R, Luccio FL, Wahsheh HAM (2018) Security threats and solutions for two-dimensional barcodes: a comparative study. In: Daimi K (eds) Computer and network security essentials, pp 207–219. Springer, Cham
8. Joo JW, Moon SY, Singh S, Park JH (2017) S-detector: an enhanced security model for detecting Smishing attack for mobile computing. Telecommun Syst 66(1):29–38
9. Sonowal G, Kuppusamy K (2020) PhiDMA–A phishing detection model with multi-filter approach. J King Saud Univ Comput Inf Sci 32(1):99–112
10. Tong W., Zheng KF, WU CH, Wang XJ (2018) SMS Phishing detection using oversampling and feature optimization method. In: Proceedings of the international conference on information, electronic and communication engineering (IECE 2018), pp 235–245. DEStech Transactions on Computer Science and Engineering
11. Sisodia DS, Mahapatra S, Sharma A (2020) Automated SMS classification and spam analysis using topic modeling. In: Proceedings of the 2nd international conference on data, engineering and applications (IDEA), pp 1–6. IEEE
12. Murdan AP, Clain DJC (2018) An android mobile application for an improved version of SMSSec, for secure SMS communication. Int J Mechatron Electr Comput Technol (IJMEC) 8 (27):3712–3724
13. GuangJun L, Nazir S, Khan HU, Haq AU (2020) Spam detection approach for secure mobile message communication using machine learning algorithms. Secur Commun Netw 1–6
14. Mishra S, Soni D (2020) Smishing detector: a security model to detect smishing through SMS content analysis and url behavior analysis. Futur Gener Comput Syst 108:803–815
15. Shahriar H, Zhang C, Dunn S, Bronte R, Sahlan A, Tarmissi K (2019) Mobile anti-phishing: approaches and challenges. Inf Secur J Glob Perspect 28(6):178–193
16. Ghourabi A, Mahmood MA, Alzubi QM (2020) A Hybrid CNN-LSTM model for SMS spam detection in Arabic and English messages. Future Internet 12(9):1–16
17. Sheikhi S, Kheirabadi M, Bazzazi A (2020) An effective model for SMS spam detection using content-based features and averaged neural network. Int J Eng 33(2):221–228
18. Jain AK, Yadav SK, Choudhary NA (2020) Novel approach to detect spam and Smishing SMS using machine learning techniques. Int J E-Services Mobile Appl (IJESMA) 12(1):21–38
19. Sonowal G, Kuppusamy K (2018) Smidca: an anti-smishing model with machine learning approach. Comput J 61(8):1143–1157
20. Apandi SH, Sallim J, Sidek RM (2020) Types of anti-phishing solutions for phishing attack. In: Proceedings of the IOP conference series: materials science and engineering, pp 1–8. IOP Publishing
21. Saha I, Sarma D, Chakma RJ, Alam MN, Sultana A, Hossain S (2020) Phishing attacks detection using deep learning approach. In: Proceedings of the third international conference on smart systems and inventive technology (ICSSIT), pp 1180–1185. IEEE

22. Choudhary N, Jain AK (2017) Comparative analysis of mobile phishing detection and prevention approaches. In: Proceedings of the international conference on information and communication technology for intelligent systems, pp 349–356. Springer
23. Wahsheh HAM (2019) Secure and usable QR codes. Università Ca' Foscari Venezia
24. Wahsheh HA, Luccio FL (2020) Security and privacy of QR code applications: a comprehensive study, general guidelines solutions. Information 11(4):1–23
25. Wahsheh HA, Luccio FL (2019) Evaluating security, privacy and usability features of QR code readers. In: Proceedings of the international conference on information systems security and privacy (ICISSP), pp 266–273. INSTICC, Czech Republic
26. Phishtank API. https://www.phishtank.com, last accessed 2021/01/15
27. Marchal S, Francois J, State R, Engel T (2014) Phishstorm: detecting phishing with streaming analytics. IEEE Trans Netw Serv Manage (TNSM) 11(4):458–471
28. Wahsheh HA, Al-Kabi MN, Alsmadi IM (2013) A link and content hybrid approach for Arabic web spam detection. Int J Intell Syst Appl (IJISA) 5(1):30–43
29. Wahsheh HA, Al-kabi MN, Alsmadi IM (2012) Evaluating Arabic spam classifiers using link analysis. In: Proceedings of the 3rd international conference on information and communication systems, pp 1–5. ACM, Jordan
30. Al-Kabi MN, Wahsheh HA, Alsmadi IM (2014) OLAWSDS: an online Arabic web spam detection system. Int J Adv Comput Sci Appl 5(2):105–110
31. Wahsheh HA, Alsmadi IM, Al-Kabi MN (2013) The evaluation of trust and credibility metrics: Websites of Jordanian universities and e-government portals as a case study. In: Proceedings of the IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT), pp 1–6. IEEE, Jordan
32. Bitly URL Shortener. https://bitly.com, last accessed 2021/01/11
33. TinyURL. https://tinyurl.com, last accessed 2021/01/11
34. Tan CL: Phishing Dataset for Machine Learning: Feature Evaluation. https://data.mendeley.com/datasets/h3cgnj8hft/1, last accessed 2021/01/11
35. Kou G, Yang P, Peng Y, Xiao F, Chen Y, Alsaadi FE (2020) Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods. Appl Soft Comput 86:1–13
36. Vaitkevicius P, Marcinkevicius V (2020) Composition of ensembles of recurrent neural networks for phishing websites detection. In: Proceedings of the international Baltic conference on databases and information systems, pp 297–310. Springer
37. Giry D: Keylength. https://www.keylength.com, last accessed 2021/01/11
38. Focardi R, Luccio FL, Wahsheh HA (2018) Usable cryptographic QR codes. In: Proceedings of the IEEE international conference on industrial technology (ICIT), pp 1664–1669. IEEE, France
39. Focardi R, Luccio FL, Wahsheh HA (2019) Usable security for QR code. J Inf Secur Appl 48:1–9
40. Witten I, Frank E (2011) Data mining: practical machine learning tools and techniques, 3rd edn. Morgan Kaufmann Series in Data Management Systems
41. Zamir A, Khan HU, Iqbal T, Yousaf N, Aslam F, Anjum A, Hamdani M (2020) Phishing web site detection using diverse machine learning algorithms. Electron Libr 38(1):65–80
42. Saxena A, Sharma N, Agarwal P, Barotia R (2021) Phishing website prediction by using cuckoo search as a feature selection and random forest and BF-tree classifier as a classification method. In: Rising threats in expert applications and solutions, vol 1187, pp 765–776. Springer

URL Links Malicious Classification Towards Autonomous Threat Detection Systems

Mohammad Alsmadi, Izzat Alsmadi, and Heider A. M. Wahsheh

Abstract Cyber threat behaviors can take different forms, approaches, and goals. For threat detection systems, it is essential to monitor URLs known for previous malicious attempts. It is also vital to study attack behaviors for the ultimate goal of designing autonomous threat detection systems. We collected a large dataset of URL links annotated toward that goal, manually as benign or malicious links. Several features are collected about those links related to lexicons and structure. Two classification algorithms were employed to extract the best features to predict whether a link is malicious or benign. We were applied three preprocessing techniques, including handling missing values, dealing with outliers, and categorical variables. For two evaluated classifiers, the results' bias is avoided using proper data splitting methods. The quality of the classifiers is evaluated using classifiers' accuracy. The results indicated that Random Forest continuously reported better accuracy than the Decision Tree classifier, which could predict URL link type, whether malicious or not, with the highest accuracy with 98%.

Keywords Cyberthreats modeling · Malicious URL links · Data classification

M. Alsmadi
Lewis University, Illinois 61348, USA

I. Alsmadi (✉)
Department of Computing and Cyber Security, Texas A&M University, San Antonio, TX 78224, USA
e-mail: ialsmadi@tamusa.edu

H. A. M. Wahsheh
Department of Information Systems, College of Computer Science and Information Technology, King Faisal University, Al-Ahsa 31982, Saudi Arabia
e-mail: hwahsheh@kfu.edu.sa

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_40

1 Introduction

Cyber-attacks continue to grow in volumes and complexity. Attackers and hackers work continuously to find new and intuitive methods to attack and avoid detection or eradication. Many websites and links exist that report on periodic bases malicious IP addresses or URLs. Those malicious activities can be related to hacking, spam, Denial of Service (DoS), or being part of an attack or malware dissemination. Some threat detection systems create blacklists of such URLs (e.g., to be blocked in the network access controls). However, using such blacklists can cause several problems or challenges [1].

The number of URLs or links in this blacklist database can be huge, with millions of links. This will cause significant overhead or performance issues on security access controls. For a network with many users or internet connections, each traffic request will be tested first against the malicious list can be a severe overhead problem [2].

Many of those URLs or IP addresses may change with time due to different reasons. Attackers may do that intentionally to avoid detection. They may use spoofing methods to alter or fake their IP addresses. Additionally, those URLs and IP addresses can be sold from time to time to different users or owners [1, 2].

As an alternative to dictionary-based malicious URLs' detection, rule-based detection methods can be used to study attack patterns. While the previous process can be more accurate in known malicious URLs' territories, detection methods based on attack, standard features can work more effectively in unknown parts [3].

In this paper, we collected a large dataset of known malicious URLs. Most of those URLs were annotated manually or verified by different tools. Another large dataset of available benign URLs is also integrated with the previous one to build a classification model. We developed a tool to extract many features of those URLs. More details on the approach and results are explained in the next sections.

Paper Focus and Contributions

Toward a goal of an autonomous threat detection system that can create its own rules based on its surrounding traffic and environment, our attempt in this paper is to evaluate the best features to distinguish malicious URLs from benign ones. Followings were our observations on surveyed relevant contributions:

- Our dataset should be large to be able to come up with realistic findings. Some papers used small size datasets, and hence results were narrow-focused, especially as the real number of URLs in the Internet world is vast.
- The evaluated dataset should have a balanced number of both malicious as well as benign URLs. Results will again be narrow-focused, if the dataset is from one category or if the size from one class is much smaller than the other.
- Features about evaluated URLs should consider different attributes about those URLs. While we acknowledge that it is impossible to collect a comprehensive list of features about those URLs that can ultimately uniquely distinguish the

different URLs, it is essential to extract features from different categories (e.g., lexicons, structure, etc.). We made sure to select features from at least those three categories for all URLs in our dataset.

- We also evaluated two classification algorithms and compared them in terms of selected features and accuracy or performance metrics. It is always vital in data classification or any data analytic activity to associate accuracy or performance results with analytical data findings.

Paper Structure

The rest of this paper is organized as follows. Section 2 presents the recent related studies of detection of malicious URLs. Section 3 explores the adopted methodology. Section 4 shows the experiments, results and discusses the main findings of this study, and Sect. 5 delivers the concluding remarks.

2 Related Work

This section summarizes several related studies to the classification of URLs or IP addresses as malicious or benign.

The study of [4] investigated the perceived susceptibility of online users' to phishing attacks. A survey for college students affirmed that user's susceptibility is influenced by the difficulty of detecting and the failures of detection outcome. Results demonstrated the relationship between the detection process's difficulty and the failures of detection outcome and determining antecedents to them.

In [5], the authors proposed a swift learning-based solution to detect phishing using the website URL. This method utilized a character-level convolutional neural network (CNN). It apprehends the URL information and patterns without requesting beforehand phishing knowledge. It used the discovered pattern features for the actual URL classification. Based on experiments, the suggested model attained about 95.02% accuracy on the study dataset, it is achieved an accuracy of 98.58%, 95.46%, and 95.22% on different datasets of benchmark, which superb the current phishing URL models. The study of [6] contributed to introducing an overall survey and a structural comprehension of malicious URL detection techniques utilizing machine learning. It approached the malicious URL detection techniques formulation as a machine-learning function and categorizing and studying the literature contribution that highlighted various dimensions of this issue. Besides, this study offered an inclusive and well-timed survey for a domain of different users, for the researchers of machine learning, academic engineers, and practitioners and professionals in cybersecurity.

The study of [7] introduced a multilayer framework to discover phishing (PhiDMA). The PhiDMA framework combines five layers: Auto upgrade whitelist layer, URL features layer, Lexical signature layer, String matching layer, and Accessibility Score comparison layer. A proposed prototype of the PhiDMA was implemented with a user interface that allows users with visual feebleness to use it

without any impediment. The results proved that the model has an accuracy of 92.72% in detecting phishing sites.

In [8], the authors proposed an anticipated similarity-based framework of phishing detection called VisualPhishNet. VisualPhishNet framework examines websites' profiles for detecting the phishing websites through a similarity metric that realized pages with new and different visual appearances. This study affirmed that the used method exceeds all previous similarity-based approaches of phishing detection by a considerable rate, as it is robust against a set of evasion attacks. Reference [9] aims to concentrate on developing and designing a phishing detection solution, which could leverage the global resource locator and different website content types, for instance, text, images, and frames. The algorithms of long short-term memory (LSTM) and convolutional neural network (CNN) were utilized to build a hybrid classification model called the "intelligent phishing detection system (IPDS)". After conducting an experimental analysis of the IPDS effectiveness in detecting phishing web pages and attacks, an accuracy rate of 93.28% of the model was affirmed. An average of 25 s of detection time was attained. In [10], the authors provided an inclusive survey of malicious web pages' current detection approaches. It offered a summarized review of the varied forms of the attacks of web pages. It also introduced URLs or IP addresses features. It also introduced URLs or IP addresses features that could be used to detect malicious web pages the malicious web pages efficiently. The study of [11] offered a design for studying the cognitive neuroscience of the detection of phishing provided a structure for studying phishing detection's cognitive neuroscience. It presented a new measure called the Phishing Email Suspicion Test (PEST), and proposed a computational model to quantify behavior. Using the PEST, candidates evaluate a set of non-phishing and phishing emails based on the suspicion level. In the proposed model, the behavior was quantified to 'candidates' suspicion level of emails, 'candidates' capability to differentiate between phishing and non-phishing emails, and the extent of bias in their decision to the recent past emails. In [12], the authors proposed an approach to reveal the phishing accruing on Ethereum through its transaction records' mining processes. They conducted a crawling of phishing addresses of two websites and remodeled the network of transactions. They then offered a new network containing an algorithm named trans2vec to detect the upcoming phishing identification addresses. Also, they utilized a "one-class support vector machine (SVM)" to distinguish the network nodes into the phishing or normal nodes. The results proved the efficacy of the trans2vec algorithm in the extraction of phishing features for transaction networks and demonstrated that the method of phishing detection serves successfully on Ethereum. Authors in [13] proceeded with a systematic, in-depth benchmarking analysis and assessment of the phishing features on various and comprehensive datasets. They suggested an unprecedented taxonomy of phishing features, depending on the purpose of that features. Also, they developed a new benchmarking framework named 'PhishBench'. The 'PhishBench' had been utilizing to assess and contrast the current phishing features. Besides, they employed the PhishBench to examine the scalability and robustness of phishing methods published in previous literature.

This study's finding indicated that the detection systems' performance is affected by the inconsistent nature of different phishing attacks.

3 Methodology

Although the presently employed malicious URLs countermeasures obtained high detection accuracy results, most of them are not appropriate for real-time prediction. Our adopted methodology consists of the following main stages:

- Data collection: we have collected a large dataset of 10,000 benign and malicious URLs from various sources, such as Phishtank [14], Kdnuggets [15], and [16]. According to the related studies [16–21], conventional learning methods are intended originally for balanced data sets. They intend to optimize their objective functions that usually guide to the highest overall accuracy (the degree of the number of true predictions out of all predictions addressed). Many studies [16–21] have shown that a balanced data set provides improved overall classification performance for several base classifiers compared to an imbalanced data set. Therefore, for this study, the dataset is divided equally into 5000 benign and 5000 malicious URLs.
- Extract Link features: There are dozens of URL properties employed to distinguish both benign and malicious URLs. Depending on previous studies in this field [16–21], we have chosen the most influential URL properties in predicting URLs classes. They become a result of applying several feature selection techniques. We developed a tool to extract 28 link-based attributes of those URLs according to the recommended detection features [16–21]. Table 1 presents the extracted attributes description. We use different common preprocessing techniques on the dataset, including:

  Handle missing values.
  Deal with outliers.
  Deal with categorical variables
  Evaluate two classification algorithms and compared them in terms of accuracy and performance metrics.

4 Experiments and Analysis

Three significant preprocessing steps changed some attributes in the dataset since we found missing values in some numeric features. To handle missing values, we analyzed the data, and the distribution reveals skewed to the right. This that the median and Interquartile Range (IQR) are more appropriate descriptive measurements than the mean and standard deviation. So, we adopt the median of the

Table 1 Link-based extracted attributes [16–21]

ID	Attribute	Notes	Type
1	Mailto	Check if the webpage includes (mailto) function	Binary
2	Existing-Frame	Checks if iframe or frame is used	Binary
3	Title Tag	Checks if the title tag is empty	Binary
4	External hyperlinks	The percentage of external hyperlinks in webpage	Numeric
5	Insecure Forms	Checks if the form includes a URL without HTTPS	Binary
6	External Form	Checks if the form includes an external domain	Binary
7	Form Action	Check if the form action includes a (#), or an empty string, or (javascript:true)	Categorical
8	Attractive words	The number of attractive words in URL*	Numeric
9	Brand-name**	Checks if a brand name appears in subdomains	Binary
10	Length of Hostname	The total characters in the hostname	Numeric
11	Domain Inside Sub	Checks if TLD is a part of a subdomain	Binary
12	IP Address	Checks if IP address in the hostname of URL	Binary
13	Queries	The number of queries in the URL	Numeric
14	Dashes1	The number of dashes in the hostname	Numeric
15	Dashes2	The number of dashes in the URL	Numeric
16	Dots	The number of dots in the URL	Numeric
17	And &	The number of (&) in URL	Numeric
18	Hash	The number of hash (#) in URL	Numeric
19	Percentage %	The number of (%) in URL	Numeric
20	Underscore	The number of (_) in the URL	Numeric
21	Tilde	The number of (~) in URL	Numeric
22	@	The number of (@) in the URL	Numeric
23	Depth-level	The depth of the path in the URL	Numeric
24	Length of URL	The total characters in the URL	Numeric
25	Domain-Paths	Checks if TLD is in the path of URL	Binary
26	Hyperlink-empty	The percentage of hyperlinks including empty value or self-redirect	Numeric
27	Frequent domain	Checks if the most frequent domain name in the webpage does not match the URL domain name.	Binary

*For example: (secure), (account), (login), (signin), (banking), (confirm)
**Brand name is the most frequent domain name in the webpage

numeric attribute to fill the missing values, and the null values were imputed using the median [22]. Data with only extreme outliers were also removed because, in a real-life scenario, removing all outliers is not wise. Hence, the rule was made a bit stringent by calculating outliers as 3* IQR (or more) and applying a numeric cleaner filter [23] to include most of the data and remove just the extreme outliers.

We have one categorical data type in our dataset. We convert it by combining levels-using Business Logic technique to the numeric attributes to overcome this challenge and enhance model performance [24]. In our preliminary assessments, two classifiers, Decision Trees (DT), and Random Forest (RF), are evaluated. Our dataset was divided into two groups: a test set and a training set. The training dataset 66% of the original dataset was used to build a model used to classify the URLs in the test set as either malicious or benign depending on their link-based features. We compute the quality measurements to describe the performance of each algorithm. We used True Positive (TP), False Positive (FP), False Negative (FN) to calculate the Precision (P), Recall (R) and F-Measure (F-M), as shown in formulas 1–3 [25].

$$Recall_i = \frac{TP}{TP + FN} \tag{1}$$

$$Precision_i = \frac{TP}{TP + FP} \tag{2}$$

$$F - measure = \frac{(2 \times TP)}{(2 \times TP) + FP + FN} \tag{3}$$

A Decision Tree (DT) is a common machine learning method applied with classification problems. DT creates sequential hierarchical decisions depending on the training data set, where every leaf expresses a class label, and the branches express features' conditions that guide to class labels. The evaluation results of Decision Tree (DT) yielded an accuracy of 96.5%. Table 2 shows more details of decision tree results.

The decision tree was able to detect malicious and benign URL classes at high accuracy results. It correctly classified 3281 URLs and an error rate of 3.5%, which means only 119 incorrectly classified URLs.

Random Forest is a strong machine learning method that can be applied for various tasks, such as regression and classification. It is an ensemble approach, indicating that a random forest model is built of many small decision trees, called estimators, which each produce their own predictions. The random forest model joins the predictions of the estimators to provide a more accurate prediction. Both DT and RF are able to handle missing values. Applying Random Forest (RF) achieved accuracy results of 97.9% and an error of 2.08%. Detailed results are shown in Table 3.

Table 2 Detailed decision tree results

Class	Precision	Recall	F-M
Benign	0.961	0.969	0.965
Malicious	0.969	0.961	0.965
Weighted Avg.	0.965	0.965	0.965

Table 3 Detailed random forest (RF) results

Class	Precision	Recall	F-M
Benign	0.978	0.981	0.979
Malicious	0.981	0.978	0.979
Weighted Avg.	0.979	0.979	0.979

RF was able to classify 3329 URLs correctly and fail with only 71 URLs for both malicious and benign classes. The early assessments indicated that DT and RF could perform well in the selected dataset and features out of those four different classifiers. In this section, the focus will be on the assessment of those two classifiers.

At the end of the experiments, RF and DT classifiers achieved high accuracy; they consistently scored accuracy above 96% in training and testing data. This resulted from several preprocessing stages between cleaning data, features selection, and proper data split between training and testing. Tables 2 and 3 summarizes the final accuracy for both classifiers.

DT classifier scores a perfect training data accuracy of 96.7% and a high testing data accuracy of more than 96.5%. RF classifier scores high accuracies for both training and testing data (98% in both). Results showed that the collected features predict with high accuracy whether a subject URL is malicious or not. Few features affect the target (Type) class than the rest of the features. This study's results regarding the detection of benign and malicious links are consistent with some previous studies [16] and [21]. Our study distinguished from them by handling the three preprocessing methods, missing values, dealing with outliers, and categorical variables with obtaining high accuracy results.

There is an initial challenge on how to deal with the categorical columns in the dataset. This was one of the main reasons to pick the DT classifier as the DT classifier deals with categorical variables with no need for further processing.

Random Forest classifier is also known to be one of the popular classifiers that can score high accuracy in many different domains. One of the main reasons for such ability is that the Random Forest classifier (RF) employs the classifier's ensemble techniques with the least of users' effort or manipulations.

5 Conclusion

In this paper, we focused on evaluating some URL and link related attributes that can be good indicators of whether a URL or link is malicious or not. This can serve as first stage alert mechanisms for threat detection systems. The goal is to enable a threat detection system to investigate suspect links with further security rules if such early and quick checking methods showed that such links are possibly malicious.

In most data analytics problems, reverse engineering approaches are used to learn from datasets with known target class labels. The learning stage is used to predict

records/datasets with unknown target class labels. In this research, our primary goal is to develop and evaluate a model to predict whether a URL link is malicious or not based on several features gathered about each URL link in the dataset. The assumption is that collected features about the dataset can convey enough information about the URL links in the dataset to predict the type of the link and to be able to distinguish malicious from benign links uniquely. Because of this, the research's contribution focuses on assessing certain data analytics activities and its impact on classification accuracy. It also evaluated the best classifiers that fit the selected dataset and the best settings for those classifiers. There are several challenges in this research, which in most cases, typical to similar research. Those challenges can either block the whole analytics process or negatively impact model accuracy.

The first problem this research dealt with was missing values in some features in some records. There are different ways to deal with missing values. Mean and median column values for each missing value were evaluated. Results indicated that using the median for missing values showed better overall accuracy in both classifiers. As such, the median value was used to replace the missing value. The difference between the median and mean indicates that the dataset may have an outlier's problem. Some feature values are extremely odd compared to the rest of the feature values. This was the second preprocessing activity that was employed. Different statistical values are evaluated to filter the outliers. Best results were accomplished using 3* IQR (or more) and applying the Numeric Cleaner filter to include most of the data and remove only the extreme outliers. Some features such as URL_ID are removed as they have no-variance. In other words, each URL has a unique URL ID.

To ensure no bias in model accuracy, we employed dataset split methods to split data between training and testing. Models learn from the training data and use this learned model to predict using the testing data. Best results in both classifiers (i.e., Decision Trees and Random Forest) are reported when split between training and testing 66/34. As overall results, both classifiers achieved high accuracy in the latest models with all preprocessing steps. Those accuracies were higher than 96% for both classifiers, in both training and testing data. Random Forest classifier, in particular, showed in most experiments slightly higher accuracy than Decision Tree. This is known in many domains, especially as Random Forest is a popular ensemble classifier that creates different classifier instances and combines them with soft weights based on their prediction confidence.

References

1. IBM Services. https://www.ibm.com/services/business-continuity/cyber-attack, last accessed 2020/11/12
2. Jang-Jaccard J, Nepal S (2014) A survey of emerging threats in cybersecurity. J Comput Syst Sci 80(5):973–993
3. Alsmadi I (2020) The NICE cyber security framework: cyber security management. Springer Nature

4. Chen R, Gaia J, Rao HR (2020) An examination of the effect of recent phishing encounters on phishing susceptibility. Decis Support Syst 113287
5. Aljofey A, Jiang Q, Qu Q, Huang M, Niyigena JP (2020) An effective phishing detection model based on character level convolutional neural network from URL. Electronics 9(9): 1514
6. Sahoo D, Liu C, Hoi SC (2017) Malicious URL detection using machine learning: a survey, pp 1–37. arXiv preprint arXiv:1701.07179
7. Sonowal G, Kuppusamy KS (2020) PhiDMA–A phishing detection model with multi-filter approach. J King Saud Univ Comput Inf Sci 32(1):99–112
8. Abdelnabi S, Krombholz K, Frit M (2020) VisualPhishNet: Zero-day phishing website detection by visual similarity. In: Proceedings of the 2020 ACM SIGSAC conference on computer and communications security, pp 1681–1698
9. Adebowale MA, Lwin KT, Hossain MA (2020) Intelligent phishing detection scheme using deep learning algorithms. J Enterp Inf Manag 1–20
10. Patil DR, Patil JB (2015) Survey on malicious web pages detection techniques. Int J u- e-Serv Sci Technol 8(5):195–206
11. Hakim ZM, Ebner NC, Oliveira DS, Getz SJ, Levin BE, Lin T, Lloyd K, Lai VT, Grilli MD, Wilson RC (2020) The phishing email suspicion test (PEST) a lab-based task for evaluating the cognitive mechanisms of phishing detection. Behav Res Methods 1–11
12. Wu J, Yuan Q, Lin D, You W, Chen W, Chen C, Zheng Z (2020) Who are the phishers? Phishing scam detection on ethereum via network embedding. IEEE Trans Syst Man Cybern Syst 1–11
13. El Aassal A, Baki S, Das A, Verma RM (2020) An in-depth benchmarking and evaluation of phishing detection research for security needs. IEEE Access 8:22170–22192
14. PhishTank Homepage. https://www.phishtank.com/developer_info.php, last accessed 2020/10/12
15. Using Machine Learning to Detect Malicious URLs. https://www.kdnuggets.com/2016/10/machine-learning-detect-malicious-urls.html, last accessed 2020/10/12
16. Chiew KL, Tan CL, Wong K, Yong KS, Tiong WK (2019) A new hybrid ensemble feature selection framework for machine learning-based phishing detection system. Inf Sci 484: 153–166
17. Abdelhamid N, Ayesh A, Thabtah F (2014) Phishing detection based associative classification data mining. Expert Syst Appl 41(13):5948–5959
18. Fahmy HM, Ghoneim SA (2011) PhishBlock: a hybrid anti-phishing tool. In: 2011 IEEE international conference on communications, computing and control applications (CCCA), pp 1–5
19. Mohammad RM, Thabtah F, McCluskey L (2012) An assessment of features related to phishing websites using an automated technique. In: 2012 IEEE international conference for internet technology and secured transactions, pp 492–497
20. Xiang G, Hong J, Rose CP, Cranor L (2011) Cantina+ a feature-rich machine learning framework for detecting phishing web sites. ACM Trans Inf Syst Secur (TISSEC) 14(2):1–28
21. Sahingoz OK, Buber E, Demir O, Diri B (2019) Machine learning based phishing detection from URLs. Expert Syst Appl 117:345–357
22. Summary Statistics for Skewed Distributions. https://web.ma.utexas.edu/users/mks/statmistakes/skeweddistributions.html, last accessed 2020/10/12
23. Class NumericCleaner. https://weka.sourceforge.io/doc.dev/weka/filters/unsupervised/attribute/NumericCleaner.html, last accessed 2020/10/12
24. Analyticsvidhya. https://www.analyticsvidhya.com/blog/2015/11/easy-methods-deal-categorical-variables-predictive-modeling/, last accessed 2020/10/12
25. Witten I, Frank E (2011) Data mining: practical machine learning tools and techniques, 3rd edn. Morgan Kaufmann Series in Data Management Systems

Face Recognition Using Neural Network Technique and Binary Pattern Methods

Mousa K. Walli, Ali Mohammed Kadhim, and Mohammed Wajeeh Hussein

Abstract This paper presents the act of face recognition using neural network techniques and binary pattern methods. This is aimed at devising a technique for robotic control via facial expression. The Matlab environment was used in this work to provide accurate processing of image and facial expression categorization using a neural network. By using the microcontroller board of basic stamp 2 (BS2), the wireless transmitter circuit was built in this study. Also, the receiver part was designed with a decoder and BS2 microcontroller and interfaced with MATLAB by serial port communication. The major function of BS2 is to remain for convinced typescript from MATLAB and process it to confirm the command required to be wirelessly transmitted. Through serial port to BS2, the four characters were sent by MATLAB, with one quality for every emotion detected. The gestures and facial expressions provide intuitional cues for interpersonal communication. Local Binary Pattern (LBP) features were introduced for texture analysis and recently have been introduced to represent faces in facial image analysis. The most important properties of LBP features are their computational simplicity and their tolerance to illumination changes. LBP features can be derived very fast in a single scan through the raw image and are within the low-dimensional feature space for texture analysis; they are applied as a local feature extraction method in facial expression recognition. This work introduced facial expression examination algorithms for motionless images and efficient extension to the sequence of images. The results showed that the classification accuracy achieved was 41% for static expression images and 86.31% for series of images; this is significantly higher than the accuracy of other standard image processing and recognition techniques.

M. K. Walli
Department of Information Technology, Technical College of Management, Middle Technical University, Baghdad, Iraq
e-mail: musa.wali@mtu.edu.iq

A. M. Kadhim (✉) · M. W. Hussein
Computer Techniques Engineering Department, Dijlah University College, Baghdad, Iraq
e-mail: ali.alsalihy@duc.edu.iq

M. W. Hussein
e-mail: mohammed.alameen@duc.edu.iq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_41

Keywords Face recognition · Neural network technique · Binary pattern methods

1 Introduction

Human face detection and recognition are considered one of the most active research problems to be solved in the fields of image processing, pattern recognition, and computer vision as it plays important roles in our life-related problems in many fields, such as medicine, device development, robotics and Human–Computer Interface researches. Facial expression can be utilized to help physically disabled persons by creating applications for them. The facial expression represents unlimited types of human emotion; however, only six major emotions are associated with all persons [1, 2]. These emotions could be categorized as surprise, fear, happiness, anger, and sadness [3, 4]. In different ways, the human face could express all types of emotion under holistic features, and or a set of features under analytic representation [5]. Facial recognition and detection are a vital research field nowadays that includes different processes and computer vision [6–9].

Facial expression is one of the detected emotions with up to 90% detection accuracy [10]. This technique, followed by physiological methods, provides 70–80% accuracy in bimodal and facial percentage [11, 12]. The implementation of facial detection could be approximately considered in 4 stages—face position resolve stage, facial breakthrough discovery stage, feature extraction stage, and reaction cataloging stage [13, 14]. A suitable facial file is used to obtain datasets for the training and testing phases; the dataset contains different images of human facial expression [15]. The initial period deals with face finding algorithm that implements the facemask recognition; this mostly deals with image pre-processing and regularization to remove dismissed areas. In the next stage, it is specified that our facial fiducial landmarks—eyebrow, eye, noses, and mouths, are recognized as the critical features for sentiment finding; hence, their feature points are removed to identify the conforming sentiment. These feature points are mined after the recognition of the feature region with edges using a bend points recognition algorithm. The Sobel straight edges recognition technique and the Shi Tomasa angle point discovery were used for these purposes. The inputs feature vectors were extracted from the facial features. The third stage involves the input feature vectors which are assumed as inputs to the MLP neural network; they are categorized according to the sentiment expressed by humans.

Facial appearance recognition systems are implemented in different ways and the popular ways rely on the examination of facial features whereas the others are based on language, paralanguage, and hybrid methods [16, 17]. An automated face recognition system was proposed by [18] as the PCA eigenface method; it was used for facial recognition and the experimental findings showed 86% recognition accuracy in a restricted environment and 80% recognition accuracy in an unrestricted environment. The drawbacks of the proposed system are that for reduced videos and large differences, it does not have good results [18]. Another approach

utilized the technique of strewn representation for identifying faces in a video. To detect the faces in an image, the Viola-Jones algorithm was used. All the measurements were performed on a real-time dataset and an accuracy of 75–80% was achieved. Their work, however, was done only for front face pictures [19]. The Robotics field has been developing for more than 30 years now and many researchers have contributed to face recognition techniques [20–29]. Still, the intelligence level of robots is very low compared to humans. Even though artificial intelligence has been applied in robotics, robots still find it hard to detect human emotion which is a trivial task for humans. So, this work is about designing an intelligent system for controlling robots through facial expression and communicating commands from laptops through wireless modules. This work introduced a new facial recognition framework based on neural networks and binary pattern techniques as demonstrated in Fig. 1.

2 Materials and Methods

Serial port communication was used to interface the BS2 microcontroller with Matlab while building the neural network-binary pattern-based framework for face recognition. Firstly, the use of a basic stamp to build the wireless transmitter and receiver based on radio frequency was analyzed. Then, the basic stamp module was programmed through serial ports to transmit and receive certain characters, starting

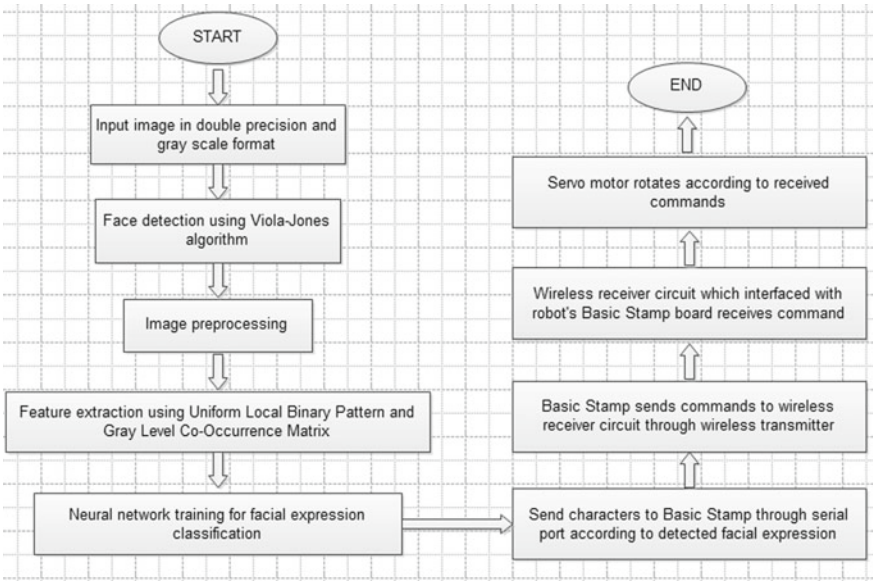


Fig. 1 Proposed face recognition techniques methodology

with a 13-bit binary rate for five seconds through the wireless system. To program the microcontroller chip, the basic stamp language was used as illustrated in Table 1. The radiofrequency of 433 MHz transmitter, encoder, BS2 kit, and the serial bus was used to confer the mentioned capability to DB9 converter cable as illustrated in Fig. 2. Figure 3 showed the pin-out feature of the PT2262 encoder used in this investigation.

The programming flow chart shown in Fig. 4 was used to program the BS2 board with basic language. The serial port communication was set to BS2 to accept the data transfer between BS2 and MATLAB. Thus, the connection speed of the serial port was set to 9600 bps with true and no parity bit conditions because a higher speed will lead to loss of data and the whole data could not be transmitted by the transmitter section. If any error occurs, it will be displayed in the LED output.

Table 1 Wireless transmitter and basic stamp operation

Facial expression detection	Characters	Transmitted 4-bit data value
Angry	"aaaaaaaaaaa"	1010
Happy	"hhhhhhhhhhh"	0101
Sad	"ddddddddddd"	0011
Surprise	"sssssssssssss"	1001

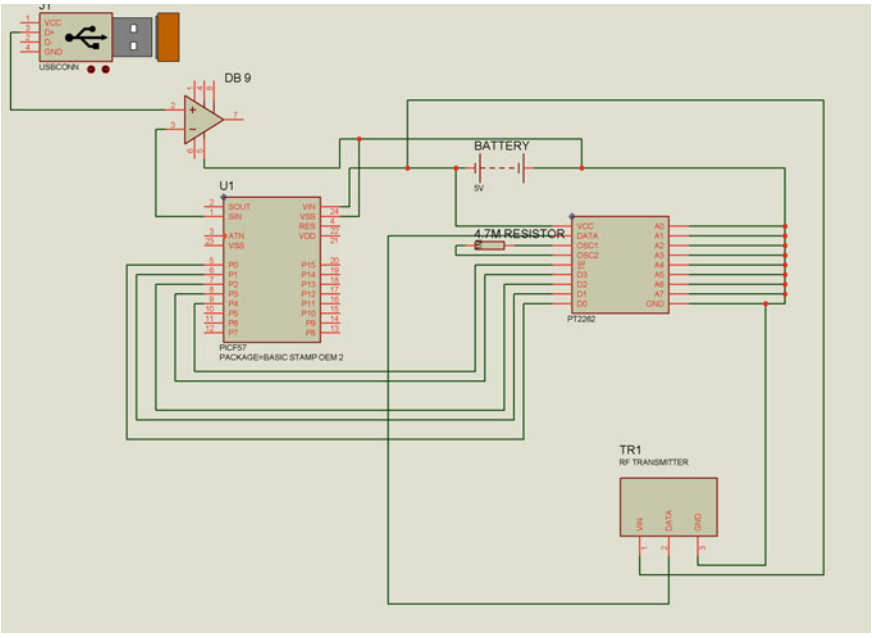


Fig. 2 Wireless transmitter system with BS2

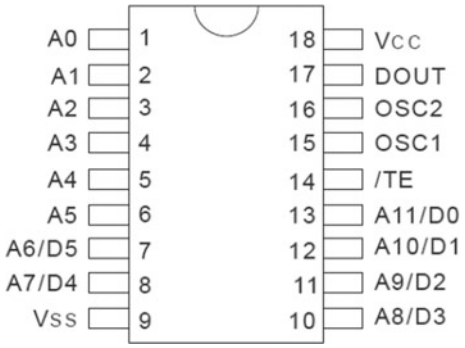


Fig. 3 Encoder output Pin

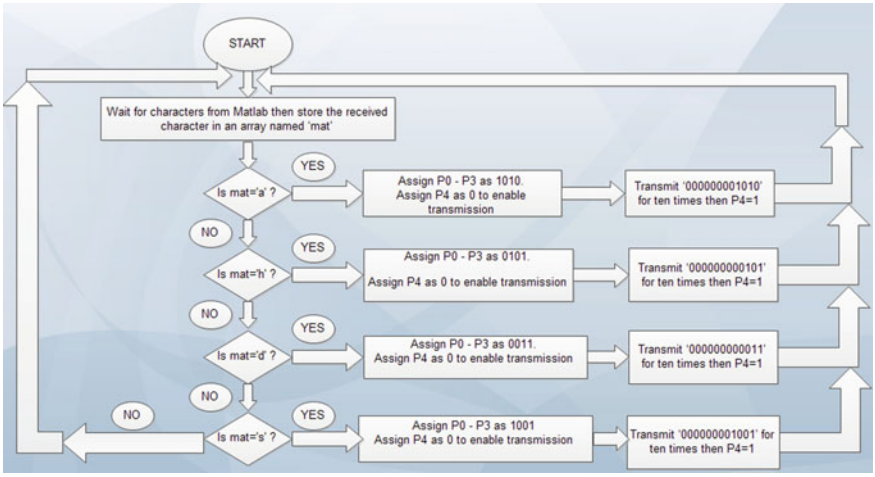


Fig. 4 BS2 board programs flow chart

The USB to DB9 converter was used to connect the BS2 board with the computer section as illustrated in Fig. 5.

After interfacing the transmitter, the receiver was developed and interfaced with the robot section to demonstrate the wireless transceiver as shown in Fig. 6. Two servo motors were used according to the received pulses from the BS2 board. The receiver section uses the PT 2272 decoder with a PIC microcontroller which is regulated by 5 V. The data sent to the BS2 board of the robot could be detected and this detection could change the rotation of the servo motors as illustrated in Fig. 7. In the case of a 1.3 ms pulse sent to the servo motor, the rotor will rotate clockwise, and this rotor will stop rotating when the 1.5 ms pulse is sent to the servo motor. Also, when the 1.7 ms pulse is sent to the motor, the rotor will rotate counter-clockwise. Hence, before the programming phase, some calculations should be

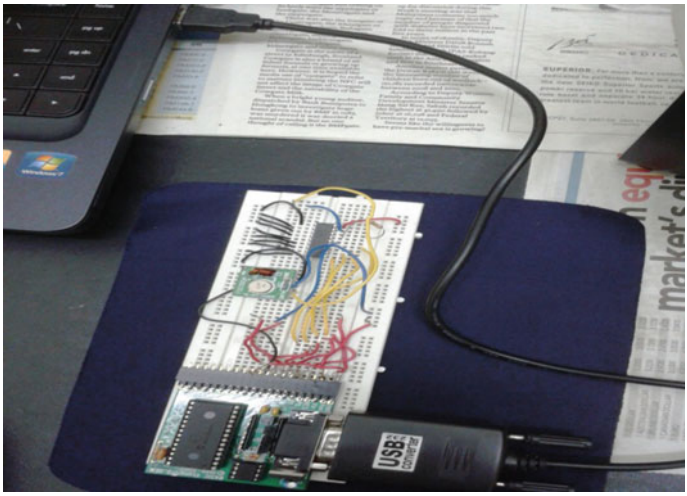


Fig. 5 BS2 to laptop Interfacing

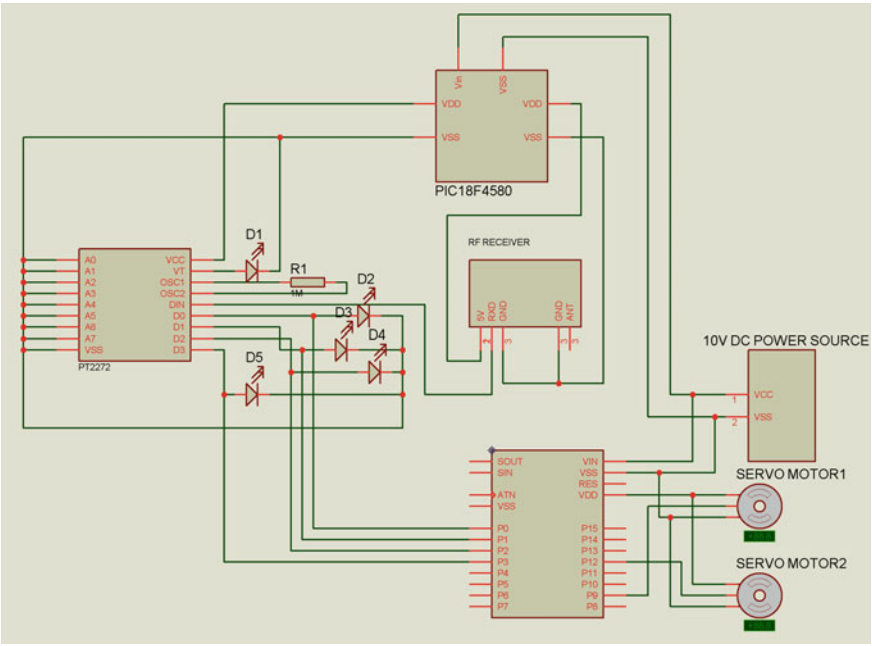


Fig. 6 Wireless receiver circuit

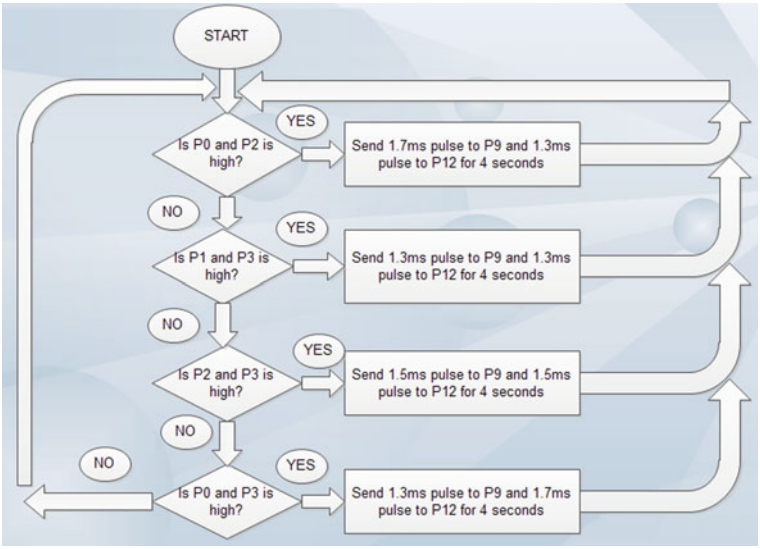


Fig. 7 Robot and BS2 programming flow chart

done to send 1.3 ms pulse via pin 9 to execute basic stamp on every command in 2 microseconds as follows:

$$\frac{1.3\,\text{ms}}{2\,\mu\text{s}} = 650 \tag{1}$$

Therefore, "PULSOUT 9,650" should be typed to launch a 1.3 ms pulse to pin 9. To state a time for high again and produce a pulse, a delay of 10 ms should be used to make pin 9 higher than 1.3 ms and go little for 10 ms to generate a pulse.

A programming loop was created to count the BS2 pulses and servo motor in 4 s continuously. Hence, to achieve these loops, an important calculation should follow to determine the number of counters. Then, to create a 4-s loop, the following code is inserted into the programs and the summation of these time values provide 23 ms and 4 s, equalling 4000 ms as follows:

$$\frac{4000\,\text{ms}}{23\,\text{ms}} = 173.9 \tag{2}$$

Then, the counter range of 173.9 should be inserted to create a loop of 4 s, and in the same way, one could calculate the BS2 values for 1.5 ms and 1.7 ms which results in 750 and 850, respectively. Table 2 showed the summary of pins that can go higher or lower than the detection limit of MATLAB for facial expression.

Depending on the received data, the LED is switched ON to compare and visualize the data between the transmitter and the receiver. Hence, in case of angry emotion being detected, then, the MATLAB program will send the characters via

Table 2 Summarize BS2 pins conditions

Facial expression Detection	Pins—high	Pins—low	Servo motor rotation 1 and 2
Angry	P0 and P2	P1 and P3	1 = counter clockwise 2 = clockwise
Happy	P1 and P3	P0 and P2	1 = clockwise 2 = clockwise
Sad	P2 and P3	P0 and P1	1 = stop 2 = stop
Surprise	P0 and P3	P1 and P2	1 = clockwise 2 = counter clockwise

serial ports to put the BS2 in the transmission mode of 1010 as shown in Fig. 8. One should be able to see here that the suitable transmission LED is elevated, showing accurate matching between the address and the other case; in this mode pin D1 and D3 can go to high levels based on the transmitted data (0101) and sad facial expression will appear on the LED output as shown in Fig. 9.

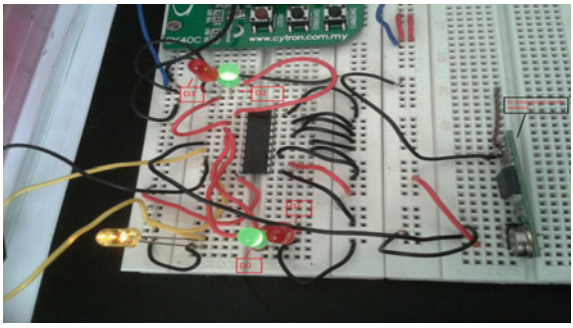


Fig. 8 Angry emotion detected by Matlab with LEDs sources

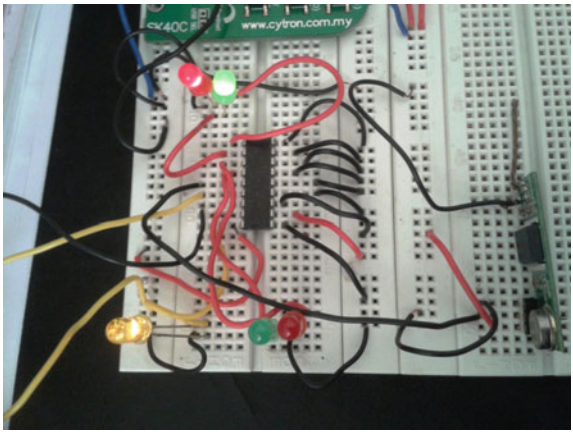


Fig. 9 Sad facial expression at LED output

Based on the data transmitted in the wireless system, the code 0011 represents surprise facial expression which is detected by MATLAB; here, D1 and D3 will turn to high levels of LED output as illustrated in Fig. 10. Then, the valid transmission happens when D0 and D3 go to the high levels due to the transmitted code, 1001. In MATLAB, the GUI was developed for simulation works using classification and image processing techniques. It integrates a call-back function in all the edit texts and pushbuttons. A fast snapshot of the call-back function creator window is illustrated in Fig. 11. The GUI test using an image was performed after the call-back function for the push button as shown in Fig. 12.

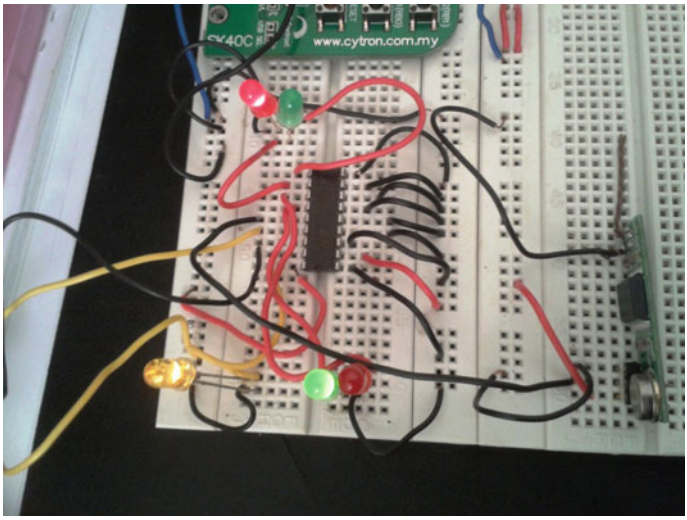


Fig. 10 Surprise facial expression in LED output

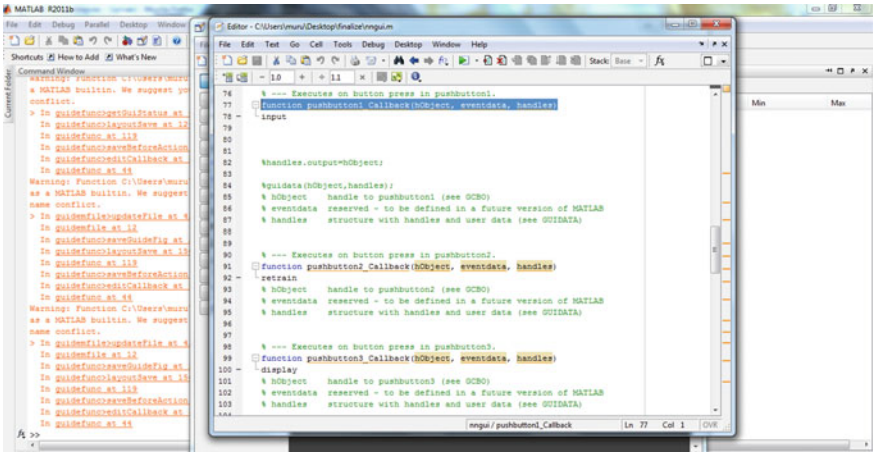


Fig. 11 Call back function window

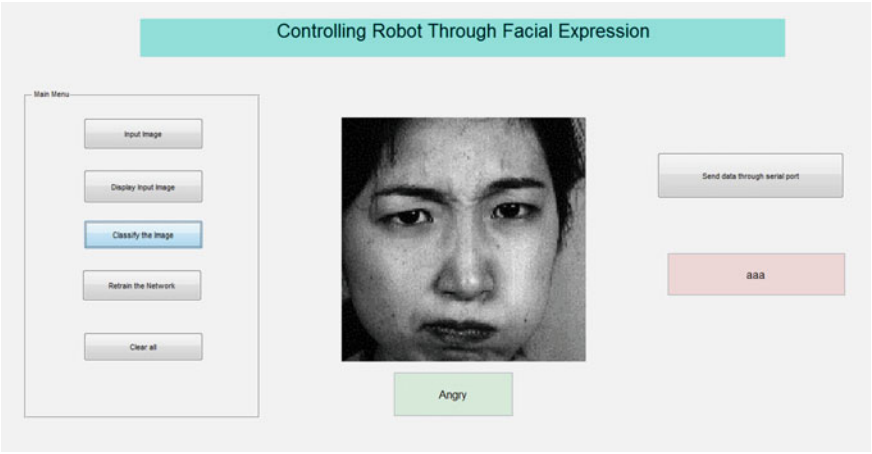


Fig. 12 Test image in GUI tester

3 Results and Discussion

Significant results were achieved with the Viola-Jones algorithm in MATLAB for face detection techniques. The image sample shown in Fig. 13 was detected based on the implementation results in MATLAB shown in Table 3 in terms of data allocation results and accuracy. The achieved test results were 58.755 under 80/20 allocation of data which explains that these results are enhanced data sets with 60/40 distribution and so on. Many trials were conducted to confirm the suggested techniques and all the achieved results are illustrated in Table 4. The adopted technique achieved 41% facial expression detection in the data set test.

As the number of epochs increases, the accuracy of the testing data also increased; however, when the number of epochs reached 4000, the accuracy

Fig. 13 Face detection as sample images

Table 3 Testing accuracy and data allocation results

Seeding weight	Allocation of data set			
	60–40%	70–30%	80–20%	90–10%
1	45.83	58.33	70.83	66.67
2	54.17	58.33	62.5	58.33
3	50	47.22	41.67	50
4	56.2	52.78	54.17	50
5	45.83	41.67	37.5	33.33
6	50	52.77	58.33	50
7	64.58	58.33	70.83	50
8	60.42	66.67	62.5	75
9	62.5	72.22	79.17	75
10	39.58	50	50	33.33
Mean	52.911	55.832	58.75	54.166

Table 4 Optimal number of hidden neuron results

Hidden neuron numbers	Testing data accuracy
31	79.17
32	70.83
33	58.33
34	62.5
35	83.33
36	54.17
37	66.67
38	83.33
39	70.83
40	75
41	58.33
42	66.67
43	54.17
44	66.67
45	62.5
46	66.67

decreased due to over-fitting. We also can see that accuracy peaked at 70.83% for 700–3000 epochs. It also showed similar behavior for 300–600 epochs as the test data accuracy remained at 66.67%. So, to obtain the best number of epochs, further experiments were initialized with a different number of weight seeds. The number of epochs was set to 700–1000. These tests were carried out with the dataset allocation of 80% (training) and 20% (testing). So, to verify this data allocation as optimum for our work, further experiments were carried out with different data allocations. Table 4 showed the determined optimal number of hidden neurons for

the neural network model (which is from 30 to 50 for our trial and error method); it can be noticed that the best performance of NN was 86.31% when the number of hidden neurons was 35 or 38. The results showed that the classification accuracy was 41% for static expression images and 86.31% for series of images and this was significantly better compared with the performance of other standard techniques.

4 Conclusion

This paper introduced new techniques for human facial expression recognition based on neural network techniques and binary pattern methods. The most important techniques in face detection are the trivial task of ensuring accurate detection. As an aspect of the field of robotics, facial expression could be exploited to support bodily disabled persons by creating a specific application for them. For example, the electronic door that turns on when the similarity of the owner's face is detected. The system in this study was created to detect facial expressions and launch a wireless command to the robots to provide a good solution for humans. This study used many pictures from the JAFFE database that contain four facial expressions. By using a uniform local binary pattern, the images were preprocessed, and the features were extracted. The emotions of the human face were classified using a neural network and 41% accuracy was achieved on the data under test. The classification was classified by sending the characters to BS2 using a wireless transmitter via serial port communication.

References

1. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. In: IEEE transactions on affective computing—May 2015. https://doi.org/10.1109/TAFFC.2014.2386334
2. Siddiqi MH, Ali R, Khan AM (2015) Human facial expression recognition using stepwise linear discriminant analysis and hidden conditional random fields. IEEE Trans Image Proc 24:1386–1398
3. Khan RA, Meyer A, Konik H, Bouakaz S (2013) Framework for reliable, real-time facial expression recognition for low resolution images. Pattern Recognit Lett 34:1159–1168
4. Ghimire D, Jeong S, Lee J, Park SH (2017) Facial expression recognition based on local region specific features and support vector machines. Multimed Tools Appl 76:7803–7821
5. Suk M, Prabhakaran B (2014) Real-time mobile facial expression recognition system—a case study. In: proceedings of the IEEE conference on computer vision and pattern recognition workshops, Columbus, OH, USA, 24–27 June 2014, pp 132–137
6. Torre FD, Chu W-S, Xiong X, Vicente F, Ding X, Cohn J (2015) IntraFace. In: Proceedings of the IEEE international conference on automatic face and gesture recognition, Ljubljana, Slovenia, 4–8 May 2015, p 18
7. Szwoch M, Pieni M (2015) Facial emotion recognition using depth data. In: Proceedings of the 8th international conference on human system interactions, Warsaw, Poland, 25–27 June 2015, pp 271–277

8. Gunawan AAS (2015) Face expression detection on Kinect using active appearance model and fuzzy logic. Procedia Comput Sci 59:268–274
9. Shen P, Wang S, Liu Z (2013) Facial expression recognition from infrared thermal videos. Intell Auton Syst 12:323–333
10. Zhao G, Huang X, Taini M, Li SZ, Pietikinen M (2011) Facial expression recognition from near-infrared videos. Image Vis Comput 29:607–619
11. Wei W, Jia Q, Chen G (2016) Real-time facial expression recognition for affective computing based on Kinect. In: Proceedings of the IEEE 11th conference on industrial electronics and applications, Hefei, China, 5–7 June 2016, pp 161–165
12. Ito Y, Ohyama W, Wakabayashi T, Kimura F (2012) Detection of eyes by circular hough transform and histogram of gradient. In: 21st international conference on pattern recognition (ICPR 2012), Tsukuba, Japan, 11–15 November 2012
13. Gao W, Zhang X, Yang L, Liu H (2010) An improved Sobel edge detection. In: 2010 3rd international conference on computer science and information technology, Chengdu, pp 67–71
14. Yin L, Wei X, Sun Y, Wang J, Rosato MJ (2006) A 3D facial Expression database for facial behavior research. In: Proceedings of the international conference on automatic face and gesture recognition, Southampton, UK, 10–12 April 2006, pp 211–216
15. Kahou SE, Michalski V, Konda K (2015) Recurrent neural networks for emotion recognition in video. In: Proceedings of the ACM on international conference on multimodal interaction, Seattle, WA, USA, 9–13 November 2015, pp 467–474
16. Ko BC, Lee EJ, Nam JY (2016) Genetic algorithm based filter bank design for light convolutional neural network. Adv Sci Lett 22:2310–2313
17. Breuer R, Kimmel R (2017) A deep learning perspective on the origin of facial expressions. arXiv:1705.01842
18. Khan A et al (2018) Forensic video analysis: passive tracking system for automated person of interest (POI) localization. IEEE Access 6:43392–43403
19. Do TD, Nguyen VH, Kim H (2020) Real-time and robust multiple-view gender classification using gait features in video surveillance. Pattern Anal Appl 23(1):399–413
20. Mohammed M, Venkata Sai Sowmya M, Akhila Y, Naga Megana B (2019) Visual modeling of data using convolutional neural networks. Int J Eng Adv Technol 9(1):4938–4952
21. Amarendra K, Mandhala VN, Chetangupta B, Geetha Sudheshna G, Venkata Anusha V (2019) Image steganography using LSB. Int J Sci Technol Res 8(12):906–909
22. Sagar Imambi S, Vidyullatha P (2018) Explore big data and forecasting future values using univariate ARIMA model in R. Int J Eng Technol (UAE)
23. Mohammed M, Sai SN, Sandhya T, Rachana S (2020) Brain tumor image classification using convolutional neural networks. Int J Adv Sci Technol 29(05):928–934
24. Mohammad Ismail B, Harsha Vardhan V, Aditya Mounika V, Surya Padmini K (2019) An effective heart disease prediction method using artificial neural network. Int J Innov Technol Explor Eng 8(8):1529–1532
25. Pisupati S, Mohammed Ismail B (2020) Image registration method for satellite image sensing using feature based techniques. Int J Adv Trends Comput Sci Eng 9(1):490–593. ISSN 2278–3091
26. Anjum G, Bhaskara Reddy T, Mohammed Ismail B, Alam M (2020) Variable block size hybrid fractal technique for image compression. In: Proceedings IEEE 6th international conference on advanced computing and communication systems, March 2020, pp 510–515
27. Shahne R, Ismail M, Prabhu CSR (2019) Survey on deep learning techniques for prognosis and diagnosis of cancer from microarray gene expression data. J Comput Theor Nanosci 16 (12):5078–5088
28. Naga Lakshmi K, Kishore Reddy Y, Kireeti M, Swathi T, Mohammed Ismail B (2019) Design and implementation of student chat bot using AIML and LSA. Int J Innov Technol Explor Eng (IJITEE) 8(6):1742–1746
29. Wu B-F, Lin CH (2018) Adaptive feature mapping for customizing deep learning based facial expression recognition model. IEEE, pp 2169–3536

Performance Enhancements of the Indoor Localization System Based on Dynamic ToA and RSS Mechanisms

Zahraa Abbas, Faeza A. Abed, Nazar J. Alhyani, and Mahmood F. Mosleh

Abstract Localization is the process of estimating the position of targets indoors or outdoors. In indoor scenarios, there are challenges to using the Global Positioning System (GPS) according to multipath environments which is the main reason. The techniques which are widely used for indoor positioning are Time of Arrival (ToA) and Receive Signal Strength (RSS). The key issues that led to elevate errors are the Non-Line of Sight (NLoS) and multipath. Many researchers are trying to address these problems using various propose techniques they approach an accuracy depending on their proposed algorithms. In this paper, a new approach is based on dynamic mechanisms to choose specific values of the path loss exponent (γ) and Gaussian variable ($X\sigma$) to match the specific environment that straight each transmitter and receiver individually based on ToA and RSS. The idea depends on taking into account the location of each recipient, which differs from the other in terms of the number of barriers that separate him and the sender. To apply the proposed system, a case study building is chosen for such a task. Also, the Wireless Insite Software (WIS) was used to simulate such building and evaluate the proposed algorithm. In such a proposal, 30 Receivers (RXs), and 3Transmitters (TXs) were deployed in suitable locations through the case study. The result obtained by WIS is supported by a special algorithm implemented by the MATLAB package to achieve the average error. The results show that the proposed scheme improves the accuracy which does not exceed (0.331) m and (0.251) m based on RSS and ToA measurement respectively in the distance above (25) m. The ToA with the proposed technique is slightly outperformed RSS in distances above (25) m and specific in the Line of Sight (LoS) scenario.

Z. Abbas · M. F. Mosleh
Electrical Engineering Technical College, Middle Technical University, Baghdad, Iraq

F. A. Abed
Institute of Technology, Middle Technical University, Baghdad, Iraq

N. J. Alhyani (✉)
Computer Engineering Techniques Department, Iraqi Ministry Of Higher Education And Scientific Research, Dijlah University College, Baghdad, Iraq
e-mail: nazar.jabar@duc.edu.iq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_42

Keywords Indoor localization · RSS · ToA · WIS

1 Introduction

The continual advancement in technology for connectivity. The advent of smart network devices has opened up several points of view to exploit the available radio signal [1, 2]. Localization-based services [1] have been one of the most significant opportunities for the use of the wireless signal for various purposes, and GPS is commonly used. Nevertheless, it only shows effectiveness in outdoor environments [3, 4], where it is successful For indoor scenarios, it can not be used effectively [5]. Because of the difficulty of Indoor Localization, there are numerous difficulties in the indoor area [6]. Interference with signs, moving vehicles, barriers, and various construction materials will sum up these issues. Over the past few years, the demand for indoor localization has grown exponentially. The development of services dependent on Indoor Localization is constant [7]. In order to monitor, operate, and estimate the propagation characteristics of the Radio Frequency (RF) network, the location of the Access Point (AP) on wireless networks is important [8]. The need for reliable and low-cost methods of localization, Many interesting ones have been opened, particularly for indoor communication, the recent forecast for researchers and industrialists [8, 9]. One approach to developing localization has been to use more complex data processing methods. Several authors have concentrated on the issue of localization and associated accuracy, where researchers have increased accuracy by gaining information based on the Angle of Arrival (AOA) [1, 10]. Even so, such techniques would take increased costs and time. On the other side, with impressive success, time-based methods like TOA were suggested in [11]. These methods, though, include synchronization of time and additional techniques to deal with timing problems. Wireless Local Area Network (WLAN) based localization was the most interesting method, as most of the buildings are fitted with Wi-Fi [12]. This focus was based on the accessibility of many wireless devices to measure signal strength results. Furthermore, it doesn't, Compared to other approaches, increased implementation costs are required [13]. As a result, several algorithms have been suggested for this reason, and the most common indoor algorithms have been based on the IEEE 802.11 standard [1, 14]. Most of these algorithms have been focused on fingerprinting methods and RSS measurements [15]. The researchers suggested an algorithm to adapt the RSS fingerprint database, for example, the researchers in [16], on the other hand, approximate the position of the AP using an algorithm based on RSS calculation, without using a database. The researchers in [17] propose RSS model algorithm to measure the various effects of noise and weak signals. On the other hand, new algorithms and approaches for AP localization without calculating reference positions have been proposed by other researchers. The most used approaches were based on the use of the Genetic Algorithm (GA), which was easily used for the problem of AP placement. In this paper, both TOA and RSS methods will use with

the dynamic system to estimate accurate locations. The aim of this research is to mitigate the challenges in a big building like NLOS and regressive signals by taking into account the impact of each environment on signal measurements.

2 Related Work

Based on Wi-Fi technologies, there are many Indoor Localization networks designed. In [18], the authors researched Indoor Localization System (ILS) under particular conditions. The key concern is the behavior of propagation in interior conditions. In comparison to Bayesian, ML was used as a mathematical approach to resolve the regressive calculations. Two algorithms were proposed to minimize the effect of NLoS. In the Bayesian approach, observations suggest that the results are more reliable than the ML estimator. In [19], the authors measured ILS based on Wi-Fi infrastructure. Subsequently, they suggested a new algorithm to resolve the effect of NLoS and improve machine accuracy. They applied their algorithms based on a few areas in order to use indoor positioning based on both ToA and RSS. The definition of two chosen areas is to obtain accurate measurements from each selected geometric area. The localization method will be carried out by substituting the enhancement data with the Triangle Centroid (TC) algorithm. The findings suggest the higher efficiency and the range of errors was smaller than another algorithm. The authors in [20] applied the ToA-based ILS model. In their study, in addition to IEEE 802.15.4 technology, they used a 2.4 GHz frequency. The authors planned to create a reliable framework by using proposed approaches and technology. Both the LoS and NLoS signal transmission possibilities have been taken into account. The findings obtained in experimental tests indicate that the UWB prefers a gap below 11 m in the LoS case. However, at distances above 11 m, 2.4 GHz is recommended. UWB is ideal for distances below 10 m in the NLOS case, while 2.4 GHz is desired over that distance. The authors in [21] presented ILS as a preferred indoor wireless networking based on Wi-Fi technology. Wi-Fi technology is spreading and promoting the use of this kind of technology by developers. The researchers benefited from APs that were used to estimate positions based on RSS methods within the building. To obtain equilibrium, the ROC curve was developed to solve the issue of dimensionality. They also do some modifying the discriminant function of the second order in order to improve simplicity. According to their findings, all those procedures led to more accuracy. In [22], the authors present ILS on the basis of a Wi-Fi approach. Deep learning is one of many techniques to enhance localization, particularly in environments indoors. Basically, there is an online phase in localization in addition to the offline phase. A database is created based on RSS measurements in the offline phase. In addition, they generated fingerprint data based on multiple algorithms to mean squared, cross-entropy data. The algorithm compares the actual measurements with the database to follow those measurements in the online process. The optimal measurements obtained after

database upgrades indicate a substantial reduction in error, so the method would be more reliable relative to previous work based on the proposed algorithm.

3 Indoor Localization

To introduce the Indoor Localization scheme, many methods are used; ToA and RSSI [23, 24] are commonly accepted methods. The exact distance is as follows, according to [25]:

$$d_i = \|X - X_i\| = \sqrt{(X - X_i)^2 + (Y - Y_i)^2} \quad i - 1.2. \ldots .k \tag{1}$$

where Y_i and X_i are the target coordinate, X and Y are the reference coordinate. Assuming one transmitter affects the receiver.

3.1 ToA Measurement

The distance to the target from the measurement device is directly proportional to the time of propagation. In order to facilitate 2-D placement, ToA measurements must be made with respect to the signals from at least the reference points. For ToA-based systems, the one-way propagation distance is estimated as [25]:

$$d_i = (time_{send} - time_{receive}) * C + m \tag{2}$$

where C is the speed of electromagnetic waves in vacuum space (C = 3 * 108 m/s), m is the variable value.

3.2 RSS Measurement

The path loss propagation model in the RSS system will be provided to calculate the distance as follows [26]:

$$PL(d) = PL_0 + 10\gamma log_{10}\left(\frac{d}{d_0}\right) + X\sigma \tag{3}$$

where PL_0 is the path loss at reference distance (d_0) (1 m), which can be calculated from the Friis equation or field measurements and is recommended for many types of research like [27]. Also, γ is the path loss exponent that depends on the required

propagation environment, and if barriers exist, the value becomes larger, d is the distance in meters between the coordinator and the mobile node, and $X\sigma$ is the random Gaussian variable, which is zero-mean (in decibels) variance.

The RSSI in dBm of the mobile node at the coordinator node can be calculated as follows [28]:

$$RSSI = PT - PL(d) \tag{4}$$

where PT is the transmitter's power in dBm, the RSSI at the coordinator node will be:

$$RSSI = PT - PL_0 - 10\gamma log_{10}\left(\frac{d}{d_0}\right) + X\sigma \tag{5}$$

It is possible to obtain the RSSI of the WiFi module from actual measurements and the $X\sigma$ can be obtained from curve fitting. It is also possible to calculate the distance between the node and the coordinator node according to [29]

$$d_i = d_0 10^{-\left(\frac{RSSI - PT + pl_0 - X\sigma}{10\gamma}\right)} \tag{6}$$

The values obtained from PL_0 in this work are 35 dBm (provided by the Friis equation for 2.4 GHz) [28], the γ depends on the setting and the range (2–6) for indoor environments, and the $X\sigma$ will take the range (2–14) [30]. In this paper, we propose to use variables $X\sigma$ and γ according to the number of wall barriers that separate between transmitter and receiver individually. For example, as shown in Fig. 2 the 3 RXs in the 1st room there is one barrier wall between TX_1 and 3 RXs in such room. On the other side, the 3 RXs in the 2nd room there are three barrier walls between the same TX and 3RXs located in the 2nd room. According to those two cases, the values of each $X\sigma$ and γ will be different between the first and second cases. Accordingly, we proposed in this paper to design an adynamic system that takes a different value for $X\sigma$ & γ to each RX depending on its location and the number of a separated wall.

4 Case Study Area

The Electrical Engineering Technical College in the Middle Technical University is selected for modeling the case study area by using WIS. Such a case study building as shown in Fig. 1 consists of 3 floors and the 2nd one selected to deploy 30 RXs and 3 TXs. The height of TXs and RXs were 2 m and 1 m respectively. The deployment of TXs was in a suitable location as shown in Fig. 2 to get convenient coverage.

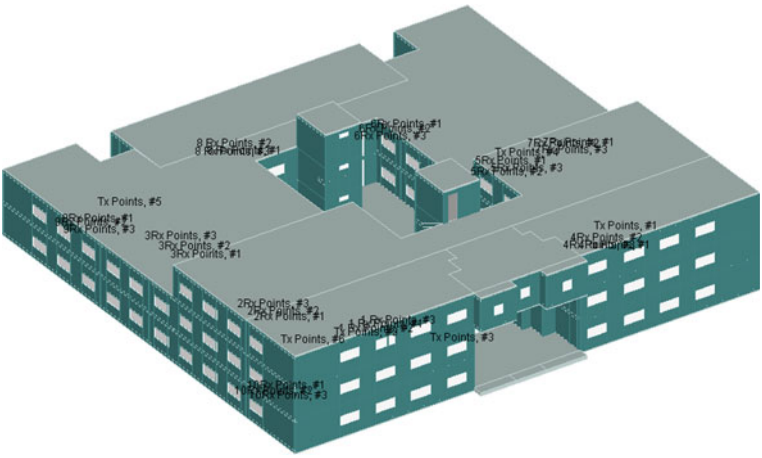


Fig. 1 The case study area

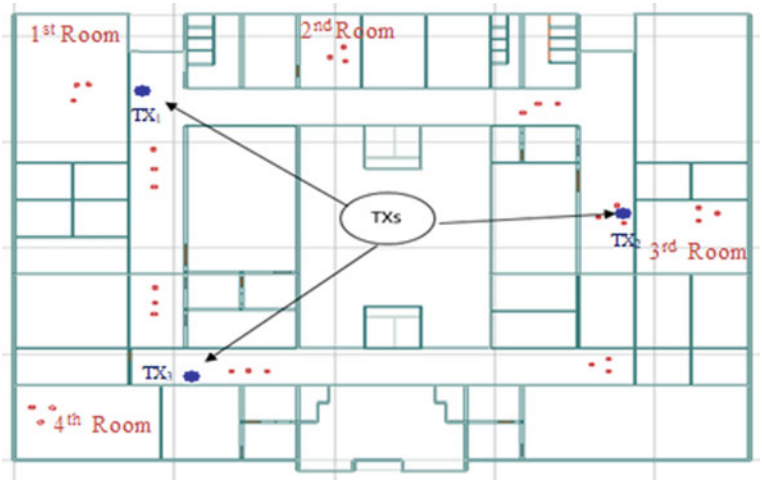


Fig. 2 The deployment of transmitters and receivers

The type of antenna was omnidirectional with a frequency spectrum 2.4 GHz and can see all characteristics in Table 1 to create a simulation model deal with real case study building, the WIS takes into account the relative permittivity (η) and conductivity (σ) for each material as recommended by International Telecommunication Union (ITU) as shown in Table 2 [31].

In Table 1 there is the capacity of WIS to set the necessary parameter to manage the individual antennas. This parameter is chosen depending on the transmitters we used so that the input power is 12.5 dBm in addition to the Sinusoid wave.

Table 1 The properties of antennas

Antenna properties	TX	RX
Antennas type	Omni.D	Omni.D
Input Power (dBm)	12.5	–
Gain (dBi)	7.5	2.2
E-Plane HPBW	13°	94°
Waveform	Sinusoid	Sinusoid
Temp. (k)	296	296
Polarizations	V	V
Received Threshold (dBm)	−150	−150

Table 2 The calculations of relative permittivity and conductivity for 2.4 GHz

Material type	Σ	η
Concrete floors	0.0931	5.21
Brick walls	0.1079	3.75
Wood doors	0.0136	1.99
Glass windows	0.0121	

5 Results and Discussions

The simulation measurements are performed in several environments into case study building. The three TXs are installed in selected locations and by using WIS, the RSS and ToA values are measured. Such measurements are included the RSS and ToA values for every 30 RXs for the signals of 3TXs individually and saved as a database.

In order to, estimate the accurate distance between each TX and all RXs, a special algorithm based MATLAB package is used to calculate all distances between each TX and all RXs the MATLAB program (see Fig. 3).

Figure 4 shows distance error between real and estimated distances. The results showed that the error range is between (0.05–0.53) m and (0.15–0.4) m for RSS and ToA respectively.

Figure 5 shows distance error between real and estimated distances. The results showed that the error range is between (0.05–0.5) m and (0.4–0.1) m for RSS and ToA respectively.

Figure 6 shows distance error between real and estimated distances. The results showed that the error range is between (0.05–0.5) m and (0.2) m for RSS and ToA respectively.

It's clear that the ToA algorithm is more suitable and accurate than RSS. The reason is because of fluctuating measurement faced RSS resulted from multipath environment By using Eq. (6), the RSS and ToA measurements convert to distances between TXs and RXs.

```
clc;
es1={};
es2={};
es3={};
err1={};
err2={};
err3={};
esdto1={};
esdto2={};
esdto3={};
erdtoa1={};
erdtoa2={};
erdtoa3={};
for i=1: 30
     % FOR RSS
    es1{i,1}=10^-((rs1{i,1}-j{i,1}+35-11)/n1{i,1})
    es2{i,1}=10^-((rs2{i,1}-12+35-11)/n2{i,1})
    es3{i,1}=10^-((rs3{i,1}-12+35-11)/n3{i,1})
    err1{i,1}=abs (rd1{i,1}-es1{i,1})
    err2{i,1}=abs (rd2{i,1}-es2{i,1})
    err3{i,1}=abs (rd3{i,1}-es3{i,1})

    % FOR ToA
    esdto1{i,1}=(toa1{i,1}*(3*(10^8)))+c1{i,1}
     esdto2{i,1}=(toa2{i,1}*(3*(10^8)))+c2{i,1}
      esdto3{i,1}=(toa3{i,1}*(3*(10^8)))+c3{i,1}
      erdtoa1{i,1}=abs (rd1{i,1}-esdto1{i,1})
      erdtoa2{i,1}=abs (rd2{i,1}-esdto2{i,1})
      erdtoa3{i,1}=abs (rd1{i,1}-esdto3{i,1})
```

Fig. 3 MATLAB program to calculate the distance

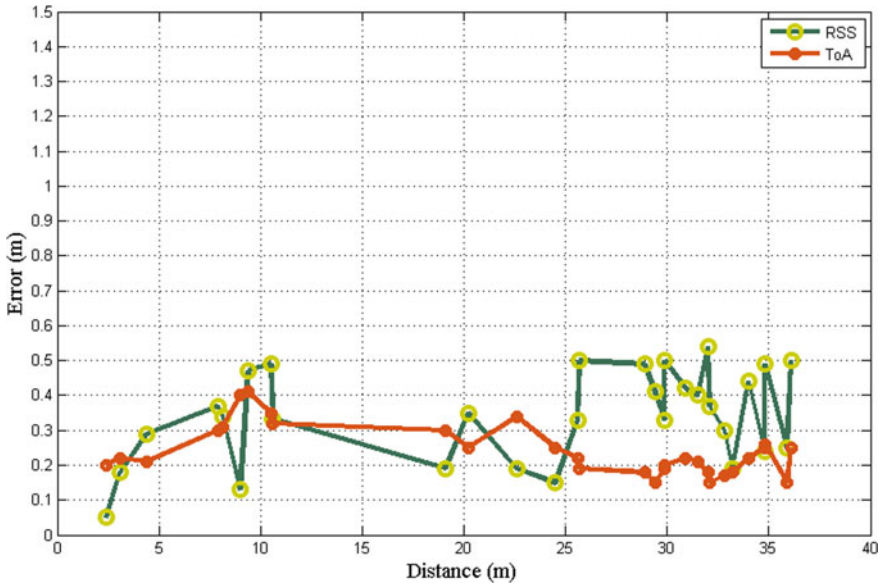


Fig. 4 The distance error in meter between the first TX and the RXs

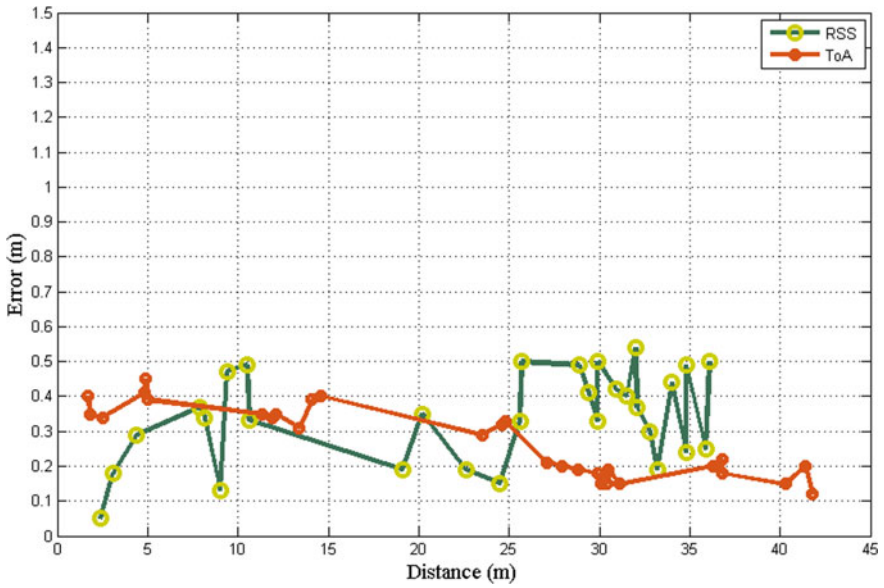


Fig. 5 The distance error in meter between the second TX and the RXs

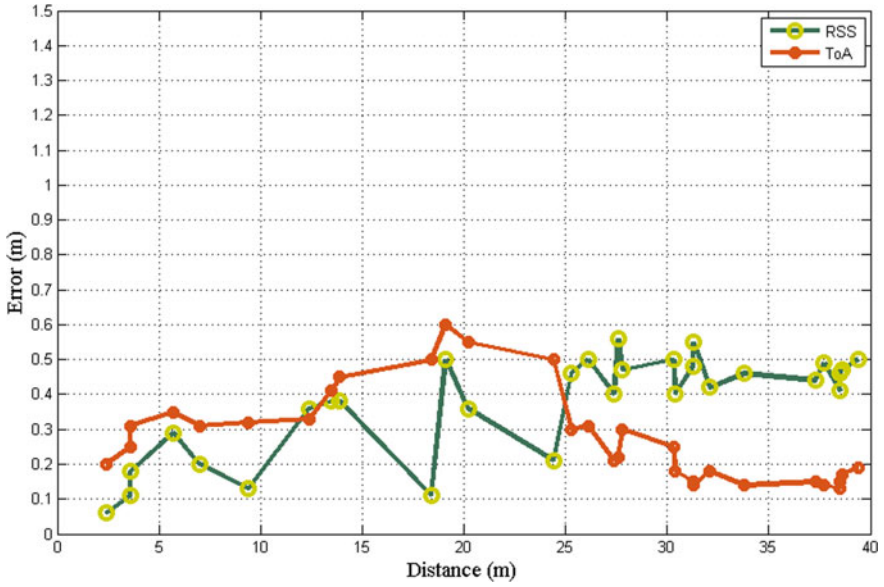


Fig. 6 The distance error in meter between the third TX and the RXs

In Fig. 4, the distances between TX and RXs were between (2.4–36.1) m and errors are better by using the ToA method in the distance exceed (25) m. In contrast, errors based on the RSS method are close to ToA methods for distances less than (25) m. As shown in Figs. 5 and 6, the errors in some points were better based on ToA rather than RSS due to LoS connections between TXs and RXs. On the whole, for all results, the average errors were (0.331) m based on the RSS method and (0.251) m based on ToA. From previous results, it can be noted that the estimated distance is close to a real distance for the distances less than (20) m and slightly increases in the distance more than (20) m, and the errors stay in the acceptable value up to (40) m. These results are compared with the research [32], where the researchers relied on static values. The results achieved in such reference reported that ToA is better than the RSS method and that errors have been identified (0.48–1.52) m. Also, the researcher of [33] used min/max positioning based on static values of both ToA and RSS measurements. The Biconical antennas were installed with the range of frequencies between (3–11) GHz, in order to apply the trilateration technique which gives better results than the min/max method with a range of errors between (0.21–2.09) m. As compared with our results which indicate that the average errors are not exceeded (0.331) m and (0.251) m based on RSS and ToA measurement respectively.

6 Conclusions

In this paper, an indoor localization system has been carried out based on both RSS and ToA methods. WIS was used for modeling a case study building that consists of three floors, in such a case study building, 30 RXs, and 3 TXs deployed. In this work, a new technique based on dynamic mechanisms based on ToA and RSS was proposed to mitigate NLoS and regressive signals. In some scenarios, the ToA method was better than the RSS method like in distances exceed (25) m with average errors (0.2) m. Whereas, in distances under (25) m, the RSS method was better with average errors (0.3) m. The ToA with the proposed technique is slightly outperformed RSS in distances above (25) m and specific in the LoS scenario.

References

1. Abdulwahid MM et al (2019) Optimal access point location algorithm based real measurement for indoor communication. In: Proceedings of the international conference on information and communication technology, pp 49–55
2. Akpakwu GA et al (2017) A survey on 5G networks for the Internet of Things: communication technologies and challenges. IEEE Access 6:3619–3647
3. Abdulwahid MM et al (2019) A comparison between different C-band and mmWave band frequencies for indoor communication. J Commun 14.10

4. Zhang J et al (2017) Path-loss-based fingerprint localization approach for location-based services in indoor environments. IEEE Access 5:13756–13769
5. Parameswaran AT, Husain MI, Upadhyaya S (2009) Is rssi a reliable parameter in sensor localization algorithms: An experimental study. In: Field failure data analysis workshop (F2DA09), vol 5, IEEE
6. Li F et al (2012) A reliable and accurate indoor localization method using phone inertial sensors. In: Proceedings of the 2012 ACM conference on ubiquitous computing
7. Hernández N, Alonso JM, Ocaña M (2017) Fuzzy classifier ensembles for hierarchical WiFi-based semantic indoor localization. Expert Syst Appl 90:394–404
8. Al-Ammar MA, Alhadhrami S, Al-Salman A, Alarif A, Al-Khalifa HS, Alnafessah A, Alsaleh M (2014) Comparative survey of indoor positioning technologies, techniques, and algorithms. In: International conference in cyberworlds (CW), pp 245–252
9. Hall PS, Hao Y (2012) Antennas and propagation for body-centric wireless communications. Artech house
10. Subramanian A, Deshpande P, Gao J, Das SR (2008) Drive-by localization of roadside WiFi networks. In: INFOCOM 2008, The 27th conference on computer communications, IEEE, pp 718–725
11. Aditya S, Molisch AF, Behairy HM (2018) A survey on the impact of multipath on wideband time-of-arrival-based localization. Proc IEEE 99:1–21
12. So J, Lee J-Y, Yoon C-H, Park H (2013) An improved location estimation method for wifi fingerprint-based indoor localization. Int J Softw Eng Its Appl 7:77–86
13. IEEE Computer Society LAN/MAN Standards Committee (2007) IEEE standard for Information technology-telecommunications and information exchange between systems-Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications. IEEE Std 802.11
14. Koweerawong C, Wipusitwarakun K, Kaemarungsi K (2013) Indoor localization improvement via adaptive RSS fingerprinting database. In: The international conference on information networking, pp 412–416
15. Abrudan TE, Haghparast A, Koivunen V (2013) Time synchronization and ranging in OFDM systems using time-reversal. IEEE Trans Inst Measur 62:3276–3290
16. Martin RK, King AS, Pennington JR, Thomas RW, Lenahan R, Lawyer C (2012) Modeling and mitigating noise and nuisance parameters in received signal strength positioning. IEEE Trans Signal Process 60:5451–5463
17. Nagy L, Farkas L (2000) Indoor base station location optimization using genetic algorithms. In: 11th IEEE International symposium on personal indoor and mobile radio communications. PIMRC,vol 2, pp 843–846
18. Al-Jazzar S, Caffery J (2002) ML and Bayesian TOA location estimators for NLOS environments. In: Proceedings IEEE 56th vehicular technology conference, vol 2. IEEE, pp 1178–1181
19. Yueyun L, Asif M (2016) A Wi-Fi-based indoor positioning algorithm with mitigating the influence of NLOS. In: 2016 8th IEEE international conference on communication software and networks (ICCSN). IEEE, pp 520–523
20. Barua B,Kandil N, Hakem N, Zaarour N (2017) Indoor localization with UWB and 2.4 GHz bands. In: 2017 IEEE international symposium on antennas and propagation & USNC/URSI national radio science meeting. IEEE, pp 1403–1404
21. Fazelinia M, Daliri MR, Ebadollahi S (2019) Wi-Fi RSS-based indoor localization using reduced features second order discriminant function. In: 2019 27th Iranian conference on electrical engineering (ICEE). IEEE, pp 921–924
22. Hsu C-S, Chen Y-S, Juang T-Y, Wu Y-T (2018) An adaptive wi-fi indoor localization scheme using deep learning. In: 2018 IEEE Asia-Pacific conference on antennas and propagation (APCAP). IEEE, pp 132–133
23. Wang Y et al (2015) RSSI-based bluetooth indoor localization. In: 11th international conference on mobile Ad-hoc and sensor networks (MSN). IEEE

24. Khudhair AA et al (2016) Wireless indoor localization systems and techniques: survey and comparative study. Indonesian J Electr Eng Comput Sci 3(2):392–409
25. Daraj AK, Mohammad AT, Mosleh MF (2020) Indoor localization enhancement based on time of arrival using sectoring method. Int J Intell Syst Appl 12(3)
26. Cho YS et al (2010) MIMO-OFDM wireless communications with MATLAB. John Wiley & Sons
27. Gharghan SK et al (2018) Adaptive neural fuzzy inference system for accurate localization of wireless sensor network in outdoor and indoor cycling applications. IEEE Access 6:38475–38489
28. Shue S, Johnson LE, Conrad JM (2017) Utilization of XBee ZigBee modules and MATLAB for RSSI localization applications. SoutheastCon 2017. IEEE
29. Ta X, Mao G, Anderson BDO (2009) On the giant component of wireless multihop networks in the presence of shadowing. IEEE Trans Veh Technol 58(9):5152–5163
30. Sahu PK, Wu EH-K, Sahoo J (2013) DuRT: dual RSSI trend based localization for wireless sensor networks. IEEE Sens J 13(8):3115–3123
31. Series P (2015) Effects of building materials and structures on radiowave propagation above about 100 MHz.“ Recommendation ITU-R, pp 2040–2041
32. Chaisang A, Promwong S (2018) ndoor localization distance error analysis with UWB wireless propagation model using positioning method. In: 2018 international conference on digital arts, media and technology (ICDAMT). IEEE
33. Monta S et al (2016) Evaluation of ultra wideband indoor localization with trilateration and min-max techniques. In: 2016 13th international conference on electrical engineering/electronics, computer, telecommunications and information technology (ECTI-CON). IEEE, pp 1–4

An RPL Based Optimal Sensors Placement in Pipeline Monitoring WSNs

Isam Wadhaj, Craig Thomson, and Baraq Ghaleb

Abstract Compared to an ordinary sensor network, Linear Sensor Networks (LSNs) has many applications in a number of areas such as surveillance and monitoring of international boundaries for illegal crossing, river environment monitoring and roads monitoring etc. However, due to linear topology and single path transmission of data, problems associated with LSNs is also significantly high. A linear topology increases the challenges associated with network reliability, connectivity and efficient energy management for sensors and actuators. It is therefore important to overcome these challenges prior to communication and reliable monitoring system that can be achieved in Wireless Sensor Networks (WSNs). This paper aims to investigate the linear placement of sensor nodes in LSNs using varying distance and its impact on energy consumption, packet delivery ratio, end to end delay, throughput and the Destination Oriented Directed Acyclic Graph (DODAG) construction time. The study is carried out using a uniform topology, where nodes are distributed linearly with increased distance, while the sink is placed at the end of the pipeline. Furthermore, the proposed study also fulfil the requirements of IPv6 Routing Protocol for Low-power and Lossy Networks (RPL) in LSNs. From the obtained results, it was evident that placing sensor nodes uniformly in LSNs at a distance of 10 m provide optimum results in terms of Quality of Service (QoS).

Keywords Internet of Things · RPL · Linear wireless sensor · Uniform node placement

I. Wadhaj (✉) · C. Thomson · B. Ghaleb
Edinburgh Napier University, 10 Colinton Road, Edinburgh E10 5DT, UK
e-mail: i.wadhaj@napier.ac.uk

C. Thomson
e-mail: C.Thomson@napier.ac.uk

B. Ghaleb
e-mail: g.Ghaleb@napier.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_43

1 Introduction

With the applications inherent in Wireless Sensor Networks (WSNs) a major driving force in development, the use of Linear Sensor Networks (LSNs) has emerged as an area of interest. An LSN becomes pertinent in situations where sensors are required to be lined up due to the particular application. This may be the monitoring of roads and bridges, however, the most obvious use is in pipelines [1–3]. This could may involve the monitoring of oil, gas or water, or sensors gathering information on the pipeline itself regarding temperature, water or oil flow, leakages, fire or pressure [1]. An added attraction of this application is in the security of transmission as opposed to cabled networks. With LSNs proven to be less prone to failure or sabotage [1].

Despite the obvious need for research in the application of LSNs, the difficulties inherent in improving these networks has resulted in a dearth of related work. As such, whilst the less complex implementation of a linear network topology may first indicate a simple implementation, this is not always the case. Whereas a complex network topology presents many different opportunities to improve the different facets of network performance, the same cannot be said for an LSN. An LSN presents little room for maneuver with regard to different combinations of Layer 3 metrics and algorithms to improve network performance in areas such as data delivery and energy consumption.

In simplistic terms, if the transmission range of any node only puts one node with-in range, as in Fig. 1, then this node will always be the next hop, whatever routing metric is utilised. This then puts a vast amount of strain on nodes nearest the sink node, which will be the recipient of every single transmission in the network. This issue can be negated by increasing the transmission range (TX) of nodes, as in Fig. 2, enabling neighbouring nodes in the linear structure to be hopped over. Whilst also bringing the possibility of utilising other metrics such as ETX or residual node energy. However, the cost of this pay-off is in increased energy consumption by each node. This occurs as radio signals are increased in order to improve said TX. This issue is highlighted in one study, which proposes the routing protocol Minimum Energy Relay Routing (MERR) for LSNs. Within this study, the authors readily admit to seeing optimal routing as the least important metric in determining the efficiency of the protocol. With the minimum use of energy seen as the highest bound [4]. In return, this also highlights another potential issue, with the previous study assuming uniform node placement. It is unlikely that this will always be the case.

The challenges inherent in improving performance levels in LSNs have been taken on in the past in different ways. The authors of one paper divided LSNs into groups based on density and the types of nodes in the network. Also seeking to inspire development of routing protocols which are more pointed in the direction of LSNs [3].

Another study attempts to negate some of the issues presented by LSNs by utilising unmanned serial vehicles (UAVs) in data collection from the sink nodes

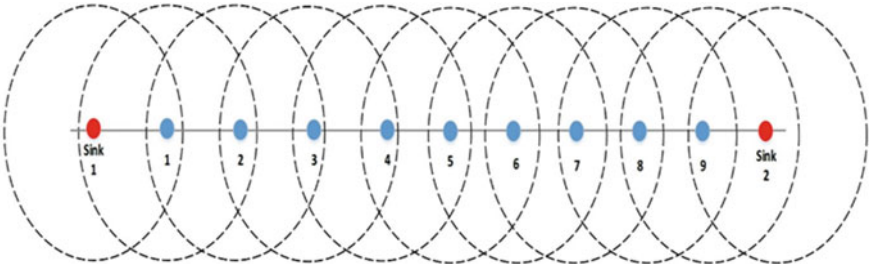


Fig. 1 Single-hop transmission range

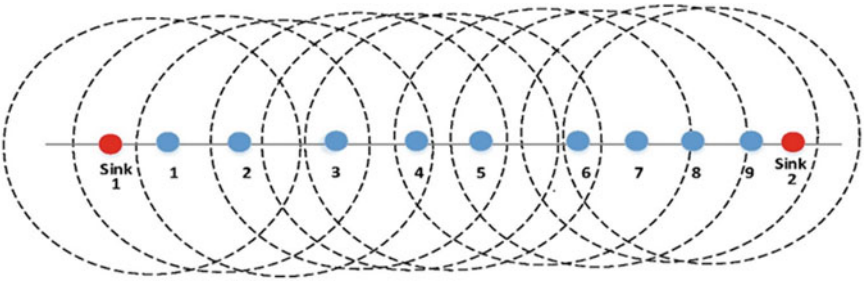


Fig. 2 Multi-hop transmission range

[2]. What is clear is that there is little consensus on how best to approach the issue of routing in LSNs.

However, given the recent standardization of the IPv6 Routing Protocol for Low-power and Lossy Networks (RPL) [5], there is a requirement to evaluate the performance of this protocol in LSNs and to improve it where possible.

The reminder of this paper is structured as follows. In Sect. 2 and, the state of art and overview about RPL is provided. In Sect. 3 related works are discussed. Section 4 presents the system model and the problem description are discussed. In Sect. 5 performance evaluation and results discussion are provided. Finally, in Sect. 6 the paper is concluded and future recommendations have been made.

2 RPL Routing Protocol Overview

In 2008 the IETF ROLL working group [6] was established with the purpose of creating a standardised routing solution for Low-power and Lossy Networks (LLNs). The result was the IPv6 Routing Protocol for Low-power and Lossy Networks (RPL) [7], a distance-vector routing protocol. RPL is, resultantly, standardised for use in WSNs such as LSNs. Considering the extreme locations in which LSNs may be utilized, the use of RPL is ideal given the allowances made for

low-power devices and unreliable infrastructure. With links expected to go down frequently and the routing protocol designed to perform accordingly.

When considering the flexibility of routing integrated into RPL, with RFC 6550 [5, 8] specifying the build of a destination oriented directed acyclic graph (DODAG), issues in consideration of LSNs become apparent. A DODAG is generally considered as a logical topology placed over a physical network. However, in the case of an LSN the logical topology is generally the same as the physical topology, therefore negating the benefit of a node being a member of multiple DODAG instances. In this case the Objective Function (OF) [9] according to which the DODAG is built becomes important, especially regarding the use of a particular routing protocol. In this case The Minimum Rank with Hysteresis Objective Function (MRHOF) [10] is of interest with its use of metrics and constraints. Although the OF itself can also be the sole driver n the DODAG build, such as in "RFC6552: Objective Function Zero" for RPL [10]. As the default OF for RPL, The Objective Function Zero (OF0) [9] does not actually utilize a routing metric. Instead using the rank of the node upon which to determine the next hop, whilst also utilizing feasible successors in the event of preferred successors not being available. This, generally, results in performance not dissimilar to when the hop-count metric is used. As such, the use of actual metrics such as Link Quality Level Reliability or ETX are recommended instead [11]. This requires the use of MRHOF [10], which unlike OF0 uses metrics to reduce the distance to a destination. These are carried in the metric container, this advertised in the DAG container option carried in DODAG Information Object (DIO) messages [8]. These used for DODAG discovery by advertising downwards in order to build routes upward towards the DODAG root, generally the sink node [11]. In the event that no metric is advertised, the default metric is ETX, the Expected Transmission Count [10]. However, many metrics are defined for use by MRHOF in "RFC6551: Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks" [11] as well as ETX, such as Hop Count, Link Latency, Node Energy or Throughput.

3 Related Work

Strategic placement of nodes is extremely crucial to satisfy contemporary performance metrics such as energy efficiency, network lifetime, coverage and connectivity. This issue is highlighted in one study, which proposes the routing protocol Minimum Energy Relay Routing (MERR) for LSNs. Within this study, the authors readily admit to seeing optimal routing as the least important metric in determining the efficiency of the protocol. With the minimum use of energy seen as the highest bound [4]. In return, this also highlights another potential issue with the previous study assuming uniform node placement. It is unlikely that this will always be the case. The challenges inherent in improving performance levels in LSNs have been taken on in the past in different ways. The authors of one study divided LSNs into

groups based on density and the types of nodes in the network. Also seeking to inspire development of routing protocols, which are more pointed, in the direction of LSNs [3].

Another study attempts to negate some of the issues presented by LSNs by utilising unmanned aerial vehicles (UAVs) in data collection from the sink nodes [2]. What is clear is that there is little consensus on how best to approach the issue of routing in LSNs. There are very few works related to node placement in LSN, they mainly focus on placement of sinks in overall LSN deployment. Whereas, focus of our work is mainly on placement of sensors because they have limited energy resources and have higher impact on network performance. Moreover, these studies do not analyse the impact of nodes location in term of quality of service metrics other than network lifetime, which we do in this work. In addition to this, the performance evaluation in this study carried out based on RPL as a routing protocol, which add more unique-ness to our work.

4 System Model and Problem Description

As a system model, a pipeline monitoring sensor network that contain N number of nodes and a sink is considered as illustrated in Fig. 3. The nodes are responsible for detecting, collecting and processing the monitored information and send it to the sink for further processing. The sink is located in one end of the pipeline.

Let N be the number of sensor nodes along the pipeline and di be the distance between node i and $(i + 1)$, $i = 1,\ldots, n$.

The length of the pipeline is

$$\sum_{i=1}^{N} d1 = L \tag{1}$$

The nodes communicate in multi-hop fashion with the same transmission range. For the data forwarding between individual nodes and the sink RPL with Expected

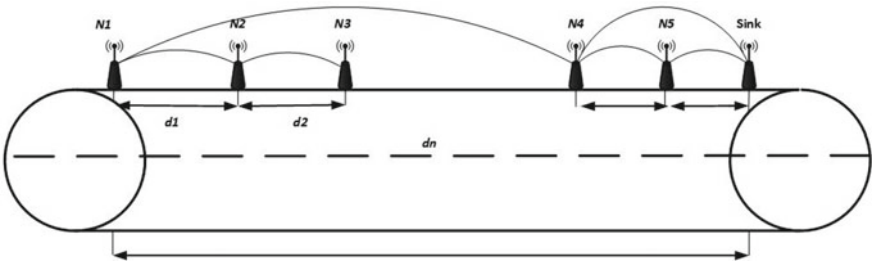


Fig. 3 Linear topology

Transmission Count (ETX) as an objective function is employed as a routing protocol. The use of the Expected Transmission Count (ETX) metric in RPL over LSN will help the nodes to use more reliable paths that is based on minimum transmissions of a packet to reach the sink.

Assuming different distance di between nodes, nodes placed at a larger distance from sink might suffer from additional overhead and overload in transmitting the packets, which in return might lead to performance implications in throughput, reliability issues and energy consumption.

5 Performance Evaluation and Discussion

In this section the simulation metrics and the performance evaluation are presented. The simulated topology consists of 24 sensor nodes and 1 sink positioned along the pipeline of length 1000 m in a linear sequential manner. The sink is positioned at the edge of the pipeline. The experiments is carried out so that an optimal spacing between nodes under uniform and linearly sensor placement schemes with increasing distance can be provided. In addition to this the performance of RPL under the same conditions can also evaluated. The evaluation has been carried out using a customized version of Omnet++ network simulator that includes an implementation of the RPL routing protocol designed by the authors. The transmission and an interference range are set to be of 100 m. The Expected Transmission Count (ETX) [12–14] is used for calculating the node ranks and selecting the preferred parent, so that stability in the network topology is built. Further parameters are provided in Table 1.

The evaluation has been carried out in terms of the following metrics:

1. The average power consumption in the network in milliwatts (Power Consumption (mW)).
2. Packet delivery ratio is the measure of the received packet by the sink over the number of packet sent by the source nodes [9, 10, 15].
3. End to end delay is the measure of the total time needed to transmit a packet over all flows in the network in millisecond [9, 10].
4. The average throughput is measure of number of packet received to the number of packet sent over the simulation time in bits per second [9].
5. DODAG construction time is the time needed for the nodes to join the DODAG.

5.1 Distance Variation

In the simulation scenario, the distance between each nodes starts at a value of 10 m then varied by 5 m to a maximum of 40 m. UDP packet of 60 bytes is sent by all

Table 1 Simulation parameters

Parameter name	Values
Simulation area	1000 × 1000 m
Number of nodes	First scenario 25
Number of nodes	Second scenario 200, 300, 400, 500
Number of packets	60 bytes
Simulation time	300 s
Mote type	Tmote Sky Mote
Mac/Adaptation layer	IEEE802.15.4/6LoWPAN
Radio model	CC2420
Transmission range (m)	30 m
Interference range	25 m
Routing protocol	RPL
Mode of operation	None storing mode
Rank metric	ETX
Nominal capacity	1000 mAh
Battery capacity	1000 mAh
Voltage	3 V
Node distribution	Linear distribution

nodes to the sink. To generate accurate results an average value of 10 runs with different seed values over a simulation duration of 300 s is taken.

Figure 4 presents the average mean power as a function of the distance between nodes. It can be observed that increasing the distance will result in an increase in the power consumption. The figure indicates that placing the node in 10 m distance from each other leads to the least amount of consumed power and the longer the distance between nodes is the more power is consumed. The power consumption reach the highest level at a distance of 40 m. This increase is due to the fact that the transmitting from a larger distance to the sink might lead to overhead and overload in delivering data packets.

Figure 5 illustrates the PDR as a function of distance between nodes along the pipeline. The delivery ratio of 98% in 10 m distance has achieved the highest level compared to the other distances followed by a delivery ratio of 92%, 89% and 87% for 15, 20, and 25 m respectively. While at distance of 30 and 35 m the PDR has dropped to 60% and 53% respectively. At a distance of 40 m, the PDR has registered the lowest value of 48% only.

Figure 6 shows the results obtained for the throughput as a function of distance in bit/s. From which it can be observed that the distance of 10 m has achieved the highest throughput of 78.72 bit/s. Surprisingly the throughput at 15 m has shown significant drop. At distance of 20, 25 and 30 m the average throughput increases outperforming 15, 35 and 40 m, but still less than the throughput in 10 m distance.

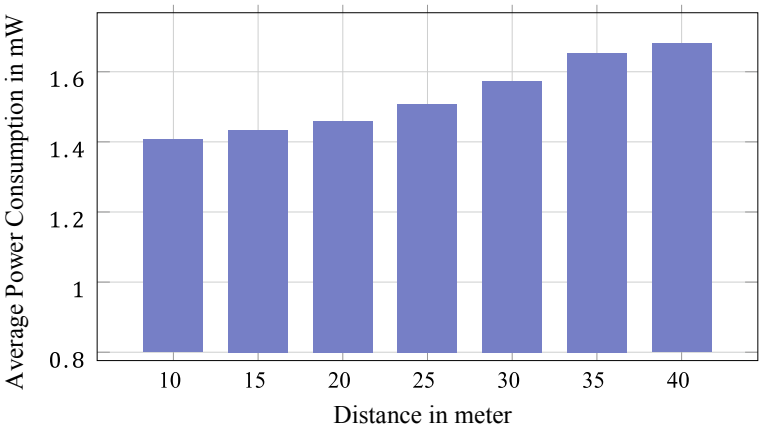


Fig. 4 The average power consumption versus distance

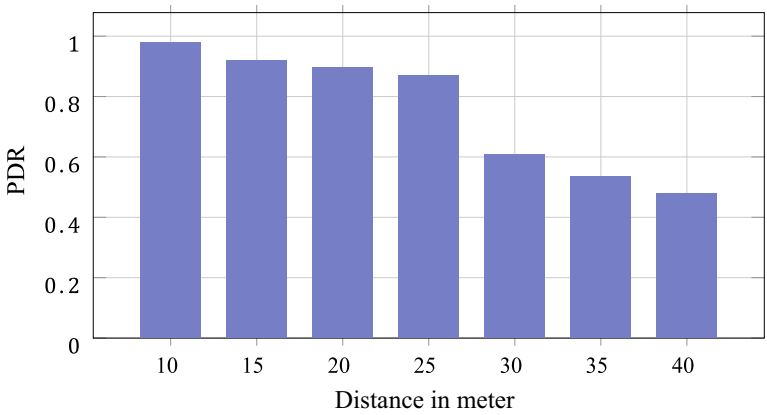


Fig. 5 The packet delivery ratio versus distance

It can be concluded that placing the node at 10 m distance from each other is the optimum in terms of throughput.

Referring to Fig. 7 the average end to end delay in packet transmission as a function of distance is presented. It can be observed that at a distance at 10 m the end to end delay has recorded the lowest level. Starting at a distance of 15 m the end to end delay increases gradually till it reaches its highest value at 40 m.

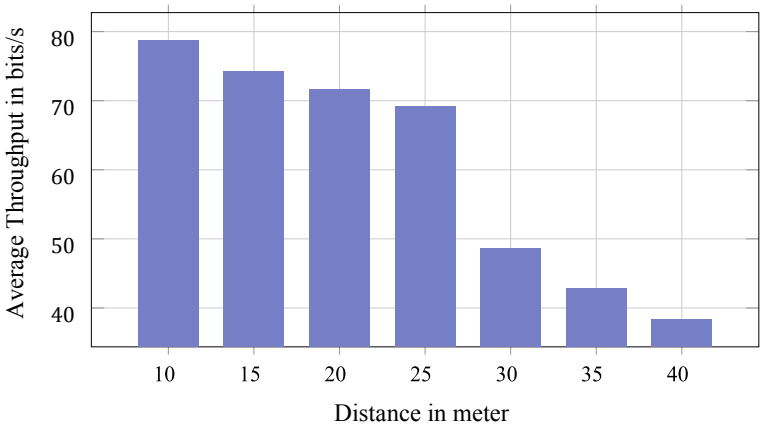


Fig. 6 The average throughput versus distance

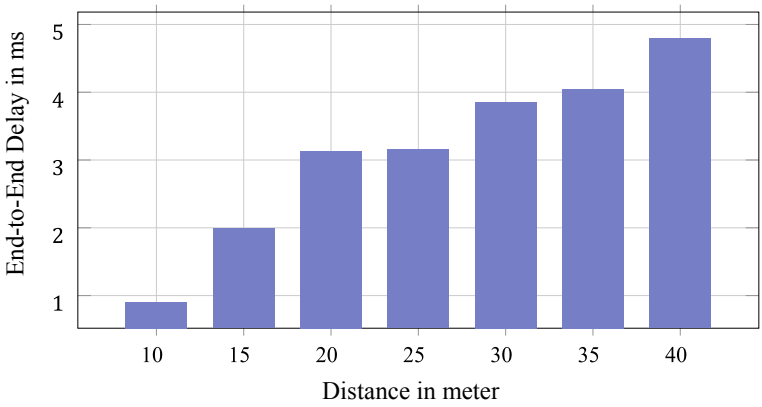


Fig. 7 The average end-to-end delay versus distance

5.2 Network Density

In the second scenario, we have increased the number of nodes starting from 100 to a maximum of 500 nodes. This required also increasing the length of the pipeline to 20,000 m to accommodate the number of nodes. The simulation time is set to 1 h to enable enough time for all node to join the DODAG. The distance between each nodes starts at a value of 10 m then varied by 5 m to a maximum of 40 m. UDP packet of 60 bytes is sent by all nodes to the sink. To generate accurate results an average value of 10 runs with different seed values. This scenario should help in studying the impact of the number of nodes and distance would affect the energy consumption, PDR, throughput and delay in a linear pipeline. Figures 8, 9, 10, 11

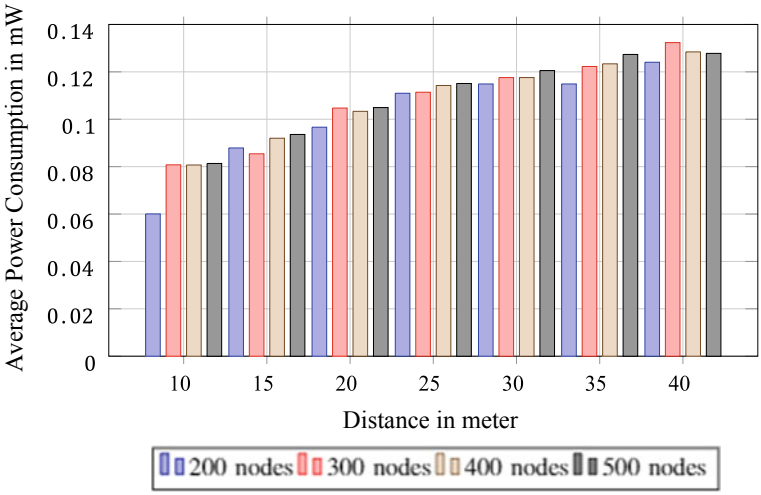


Fig. 8 The average consumption versus distance for different node number

and 12 present an inter-distance study for various number of nodes in terms of the average power consumption, PDR, End-to-End delay, and DODAG construction time.

Figure 8 represents the power consumption for different distances and increasing number of nodes. It can be observed that in a distance of 10 m the power consumption is the lowest. The power consumed is increased by increasing the distance and the number if nodes. This happens because the number of hops needed to reach the destination is increased by increasing distance, and hence the average power dissipated is decreased, but this is valid up to an optimum value, which is here the distance of 10 m. It is clear from the Fig. 12 that there is an inverse relationship between the distance and the power consumption while there is a direct relationship with the delay and DODAG construction time. Indeed, the lesser is the minimum DIO interval, the higher is the energy consumption.

From Figs. 9 and 10 the packet delivery ratio and throughput decrease with increasing the distance, and increase as the number of nodes increases. Here also we can conclude that 10 m distance is an optimal distance that can accommodate up to 500 nodes. This also can be justified by considering Fig. 12, where it shows shorter the distance the less time it takes for nodes to join the DODAG. As consequence immediate packet transmission can start.

A Poor packet delivery rate is an ingrained problem. It may be caused by many reasons, such as interferences, collisions, signal attenuation etc. The Fig. 9 illustrates the delivery ratio according to different distance and node density that are classified to their Euclidian distance from each other and from their sinks. The path cost selection is based on the OF with ETX metric, which unsurprising gives the

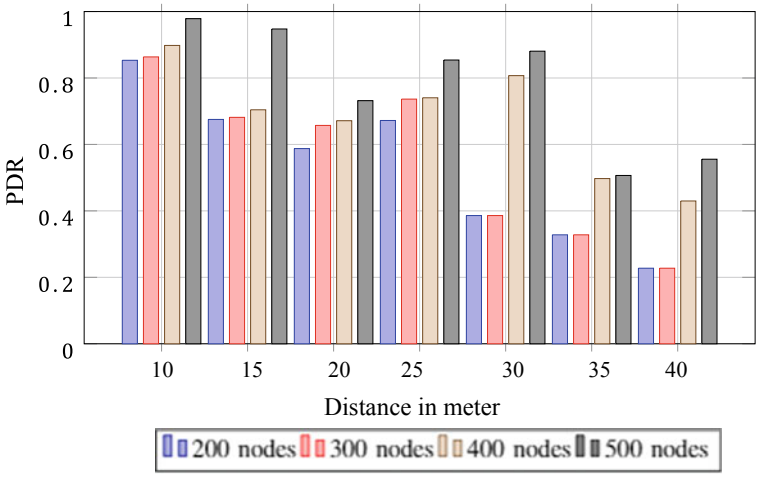


Fig. 9 The packet delivery ratio versus distance for different node number

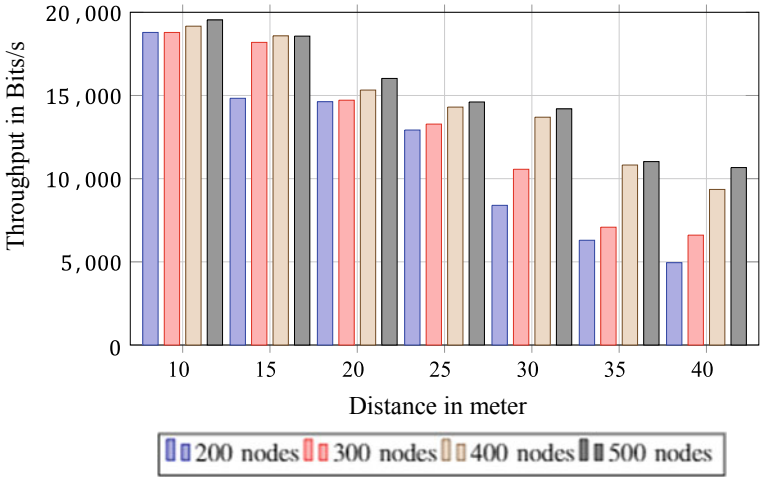


Fig. 10 The throughput versus distance for different node number

best overall results in terms of power consumption, PDR and throughput. However, this comes at the cost of increased delay due to the probing packets required to calculate the ETX metric. In addition, the results shown in Fig. 12 is applicable to justify the poor delay when the distance increases.

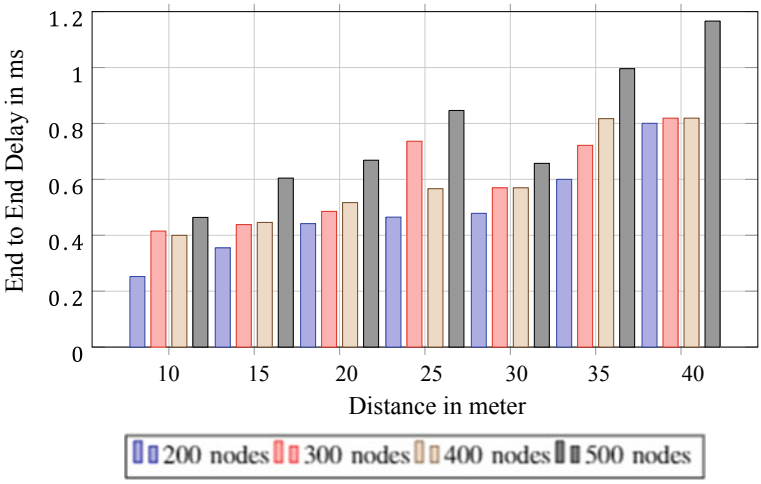


Fig. 11 The end-to-end delay versus distance for different node number

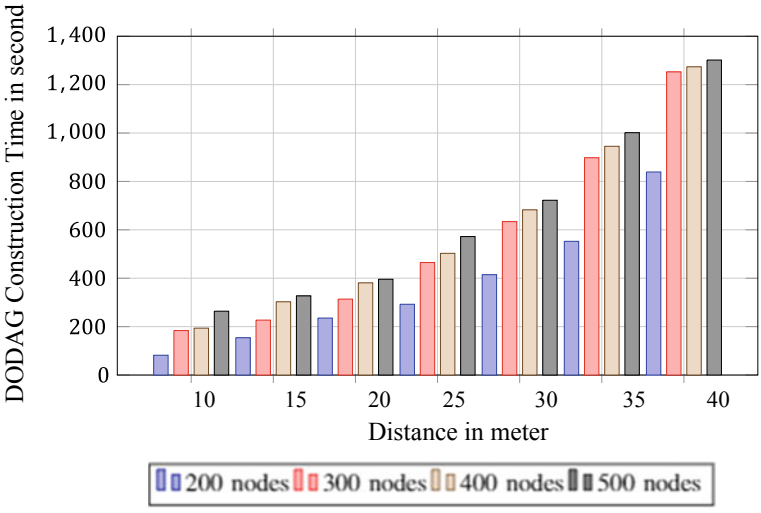


Fig. 12 DODAG construction time versus distance for different node number

6 Conclusions and Future Work

This paper has highlighted a research carried out to investigate the impact of node placement in a linear sensor network (LSN) by placing the nodes uniformly a long a pipeline. The study based on increasing the distance between nodes, while placing the one sink at the end of the pipeline. The nodes communicate in multi-hop fashion

with the same transmission range. For the data forwarding between individual nodes and the sink RPL with Expected Transmission Count (ETX) as an objective function is employed as a routing protocol. The study evaluate the performance in terms of power consumption, throughout, PDR and end to end delay. The findings have shown that the increasing of the distance between nodes has important performance implication in terms of the studied quality of service metrics. The shorter the distance between nodes the better the performance is. It can be suggested with great confidence that a distance of 10 m has achieved a significant level of performance and can be considered as the optimum solution for node placement in an LSN. In the future it is planned to look at linear network from a routing prospective in LSN by conducting further study on RPL performance compared to the traditional standards, but also within RPL itself by using different objective functions to determine the best path.

References

1. Jawhar I, Mohamed N, Shuaib K (2007) A framework for pipeline infrastructure monitoring using wireless sensor networks. In: 2007 wireless telecommunications symposium, pp 1–7, IEEE
2. Jawhar I, Mohamed N, Al-Jaroodi J, Zhang S (2013) Data communication in linear wireless sensor networks using unmanned aerial vehicles. In: 2013 international conference on unmanned aircraft systems (ICUAS), pp 492– 499, IEEE
3. Jawhar I, Mohamed N, Agrawal DP (2011) Linear wireless sensor networks: classification and applications. J Netw Comput Appl 34(5):1671–1682
4. Zimmerling M, Dargie W, Reason JM (2007) Energy-efficient routing in linear wireless sensor networks. In: 2007 IEEE international conference on mobile Adhoc and sensor systems, pp 1–3, IEEE
5. Winter T, Thubert P, Brandt A, Hui JW, Kelsey R, Levis P, Pister K, Struik R, Vasseur J-P, Alexander RK (2012) Rpl: Ipv6 routing protocol for low-power and lossy networks. rfc, vol 6550, pp 1–157
6. Ietf, "routing over low power and lossy networks (roll)." http://datatracker.ietf.org/wg/roll/charter. Accessed 2019–10–16
7. Winter T, Thubert P, Brandt A, Clausen T, Hui J, Kelsey R, Levis P, Pister K, Struik R, Vasseur J (2011) Rpl: Ipv6 routing protocol for low power and lossy networks. draft-ietf-roll-rpl-19. Technical report, Internet Draft, March
8. Dunkels A (2010) Interconnecting smart objects with IP. Morgan Kaufmann
9. Thubert P et al (2012) Objective function zero for the routing protocol for low- power and lossy networks (rpl)
10. Gnawali O, Levis P (2012) The minimum rank with hysteresis objective function. RFC 6719
11. Vasseur J, Kim M, Pister K, Dejean N, Barthel D (2012) Routing metrics used for path calculation in low-power and lossy networks. In: RFC 6551, pp 1–30, IETF
12. De Couto DS, Aguayo D, Bicket J, Morris R (2003) A high-throughput path metric for multi-hop wireless routing. In: Proceedings of the 9th annual international conference on Mobile computing and networking, pp 134–146
13. Ghaleb B et al (2019) A survey of limitations and enhancements of the IPv6 routing protocol for low-power and lossy networks: a focus on core operations. In: IEEE communications surveys and tutorials, vol 21, no 2, pp 1607–1635, Secondquarter 2019

14. Ghaleb B, Al-Dubai AY, Ekonomou E, Romdhani I, Nasser Y, Boukerche A (2018) A novel adaptive and efficient routing update scheme for low-power lossy networks in IoT. IEEE Internet Things J 5(6):5177–5189
15. Malkin G (1988) Routing information protocol version 2: request for comments: 2453

Wireless Sensor Networks (WSN) in Oil and Gas Industry: Applications, Requirements and Existing Solutions

Isam Wadhaj, Craig Thomson, and Baraq Ghaleb

Abstract Effective measurement and monitoring of certain parameters (temperature, pressure, flow etc.) is crucial for the safety and optimization of processes in the Oil and Gas Industry. Wired sensors have been extensively utilized for this purpose but are costly, not best suited for harsh environments and are difficult to deploy and maintain. Wireless Sensor Network Solutions is revolutionizing the Offshore Oil and Gas industry providing evolving solutions that introduces significant benefits in cost, ease of deployment, flexibility and convenience. The adoption of Wireless Sensor Networks is expected to be tremendous in industrial automation owing to a report that projected the deployment of 24 million wireless-enabled sensors and actuators worldwide by 2016. With limited literature on this specific subject matter, this paper presents a critical survey into oil industry monitoring specifications, requirements and Wireless Sensor Network applications as it directly impacts the Oil and Gas Industry. An overview of Wireless Sensor Networks is presented, applications from literature are highlighted and finally challenges and existing solutions are discussed.

Keywords Oil and Gas · WSN Standards · Pipeline monitoring

1 Introduction

The energy industry increasingly dependent on information technology. Although the oil and gas companies may have not invested the same percentage of money that is invested by other industrial sectors in information technology, such as the

I. Wadhaj (✉) · C. Thomson · B. Ghaleb
Edinburgh Napier University, 10 Colinton Road, Edinburgh E10 5DT, UK
e-mail: i.wadhaj@napier.ac.uk

C. Thomson
e-mail: C.Thomson@napier.ac.uk

B. Ghaleb
e-mail: g.Ghaleb@napier.ac.uk

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_44

financial services sector, but that there is a growing awareness of the role of this technology in shaping the future of the industry. Since the world economic growth depends largely on the oil and gas industry and the demand for energy resources increases, there is a need to adapt to intelligent technologies for the improvement of all areas of industrial practice in connection with the oil and gas processing stages including refining, exploration, extraction, transport and marketing of petroleum products [1] and thus increases the productivity while reducing costs. In the light of the above mentioned phases, the hash and intensive remote environments of oil and gas plants, required an adaptive monitoring system that is ideal for temporary installations, flexible, adaptable and in a position to reduce complexity and cut operational costs. In the last few years, oil and gas industries have used wired communications as a solution for applications monitoring. The installation, operation and maintenance of such a solution are usually costly and not appropriated for short-term installations and might be not be easy to adapt in a harsh environments [2, 3]. The major benefits of the sensor nodes lie not only in their small size and self-organising capabilities but also in their ability to provide reliable, fast, flexible, secure and cheap wireless communication. These features in wireless sensor nodes has introduced a convenient alternative to the wired nodes [3, 4].

The technical capability of wireless sensor networks (WSN) is still under ongoing exploration in science, as its true potential has not been fully exploited. This survey aims to provide WSN designers, oil and gas companies and researcher a critical review at the provision of WSN in the oil and gas industry taking into account the specific requirements of the applications and challenges. Furthermore, a particular attention is given to the existing architectures of efficient mechanisms that fulfil the requirements and overcome the challenges that arise when deploying a WSN in Oil and Gas industry. The paper is organized as follows. In Sect. 2, the potential applications in oil and are identified. Then, in Sect. 3, existing standards for low-power wireless sensor networks are discussed. In Sect. 4 the technical requirements of WSN deployments in oil industry are reviewed. In Sect. 5, we review techniques and solutions proposed in the literature. Finally, Sect. 6 concludes the survey.

2 Wireless Sensor Network Potential Applications in Oil Industry

Oil and Gas Industry WSNs offer a large amount of applications that are useful for the production of Performance Optimization [1] (monitor pipelines, gas detection, corrosion, H2S, equipment health [2] status, and real-time reservoir and process control, safety, maintenance. This section will present a classification of WSN possible applications in oil and gas industry Fig. 1 illustrates a classification for potential WSN applications in oil and gas industry.

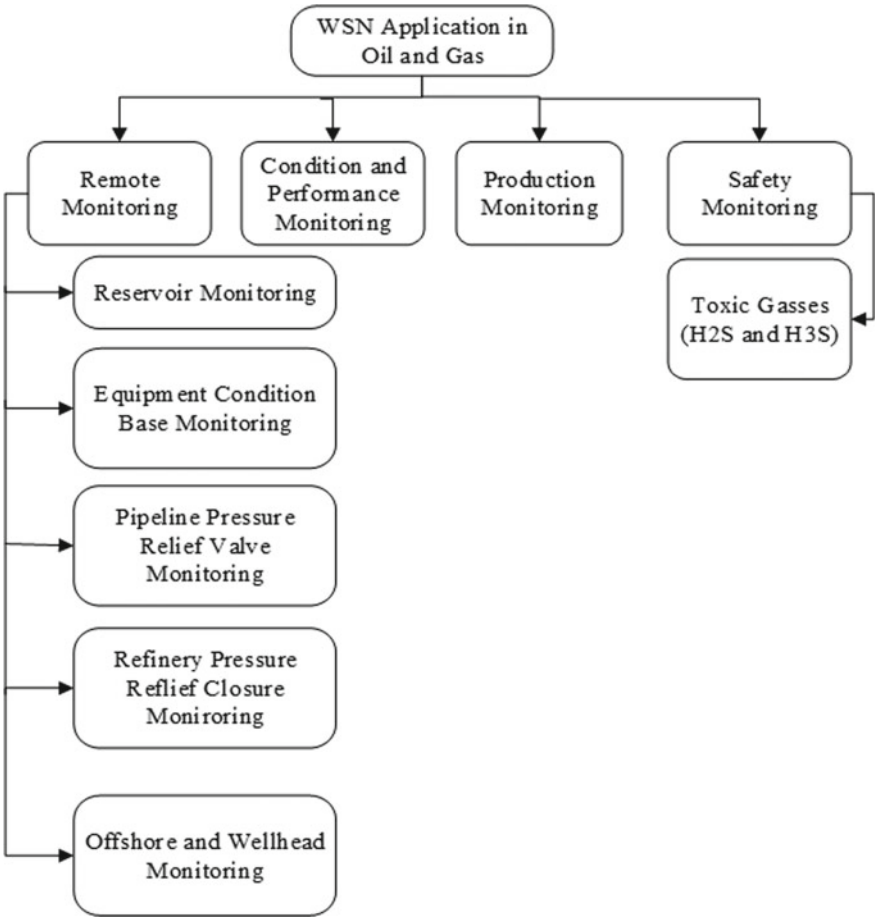


Fig. 1 Classification of WSN applications in oil and gas industry

2.1 Remote Monitoring

One of the main purposes of Remote Monitoring is to monitor Pipelines for gas leakage and damages [5]. Remote Monitoring is very beneficial for oil and gas industry in improving safety, enhancing operations, detecting problems and reducing overall operational costs [1]. Hence sensor nodes are placed in numerous remote locations and hazardous environments such as in oil and gas pipelines can suffer from a variety of catastrophic events such as explosions due to high flammability [6]. This will result in serious environmental hazard and financial damage due loss of production Therefore it is crucial for the real time monitoring system to meet certain requirements to be able to predict possible failures before they occur.

In Pipelines monitoring the unique long distance linear topology characteristics of the network infrastructure in which a large pool of sensor nodes are distributed linearly along the pipeline and limited to a single path to transmit the data, increases the challenges associated with network reliability, connectivity and an efficient energy management for sensors and actuators. The main requirements identified of WSNs applications for remote monitoring are delay, robustness, data reliability and security [7, 8].

2.2 Condition and Performance Monitoring

In oil and gas industry as in many other industries condition and monitoring of equipment and machinery plays an essential part in the overall operation. The aim of condition and monitoring is to provide fault diagnostics of different equipment and conduct machine health monitoring and temperature monitoring. Wireless sensors can be used to detect vibration, temperature heat, dissolved gas, electromagnetic properties, power consumption [1], to gather information about the health status of machineries and provide fault diagnostics, that help identify the root cause of a problem, detect and even predict potential upcoming faults, thus the operation downtime, repair costs, damage and potential danger are minimized.

2.3 Safety Monitoring

The exploration and refinery procedures of oil is usually accompanied by several toxic gases that include ammonia (NH_3), hydrogen sulphide (H_2S), and sulphur dioxide [1, 7]. The leakage of such materials could have a serious impact on human beings and the environment. Therefore, the monitoring of the H_2S considered important WSNs for prospecting for oil and gas companies [9, 10].

3 Wireless Sensor Network Standards

As WSNs and their applications become more widespread in the oil and gas industry, companies are challenged with the decision of choosing between several emerging technologies and standards, such as IEEE802.15.4, Zigbee, ISA100.11a or WirelessHART. These standards are developed with the goal to provide high-level communication protocols for WSNs and build the foundation of a complete network infrastructure taking into consideration the limited resource of the wireless sensor nodes in terms of power, reliability, security, etc.

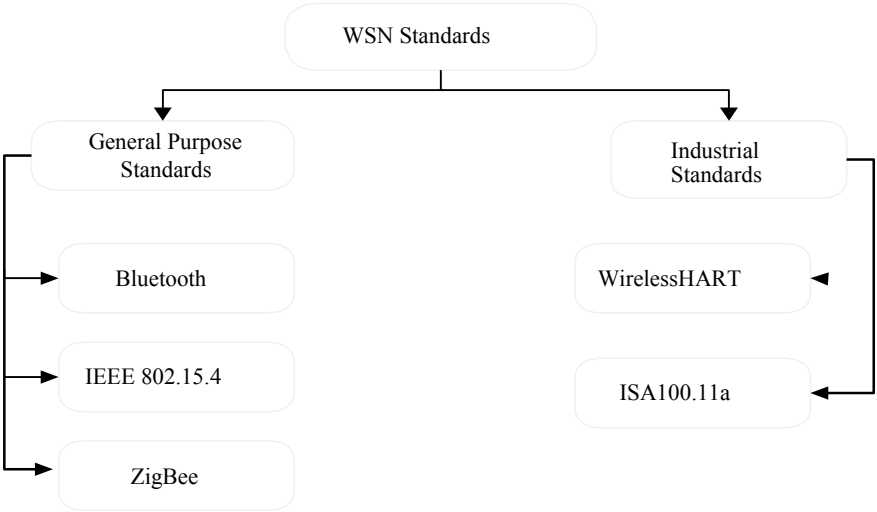


Fig. 2 Classification WSN standards

In this section, we provide an overview of the existing standards. A detailed Comparison of WirelessHART, ISA100.11a and ZigBee for industrial applications is studied in [11]. Figure 2 illustrates a classification of existing WSN standards.

3.1 A. Bluetooth

Bluetooth [12] is designed as an open wireless communication protocol to provide cost efficient services for devices that operate within short distance range with no longer than 10 m with 2400 and 2483.5 MHz frequency range. Bluetooth support the communication between 1 master and more than 7 nodes per piconet [13]. In terms of energy consumption Bluetooth has limited battery life time. However, Bluetooth can be very beneficial for a wide range of application that are deployed in a short distance range such as Personal Area Network (PAN).

3.2 IEEE 802.15.4

IEEE 802.15.4 [14, 15] specifies the physical layer and Medium Access Control (MAC) for low data rate Wireless Personal Area Networks (WPANs) LR. This standard uses a duty cycling mechanism with which all participants nodes go into a sleep state in a regular interval so that energy can be saved. The standards uses two modes: (1) non-beacon modes in which a Carrier Sense Multiple Access with Collision Avoidance (CSMA-CA) (2) a beacon mode in which the PAN

coordinator sends beacon frames to detect the PAN and starts node synchronisation. It also sends a super-frame structure that contains 16 time slots the so called guaranteed time slots (GTS) that can be assigned to nodes in each PAN, where only a maximum of seven GTS slots can be assigned to nodes contention free period (CFP). Other nodes can either transit into a contention access period (CAP) or inactive period of time during which the end-devices and coordinator transit into sleep state.

3.3 Wireless Networks for Industrial Automation—Process Automation (WIA-PA)

The Chinese Industrial Wireless Alliance [16] has designed the Wireless Networks for Industrial Automation—Process automation (WIA-PA) that is based on IEEE 802.15.4 without modifications [17] with aim to provide a reliable, energy efficient, and multihop mesh network that adapt to frequent changes in the network. The standard uses a combination of CSMA, TDMA, and FDMA with a maximum of I6 channels under the 2.4 GHz band and frequency hopping [18]. It uses three types of frequency hopping mechanisms namely the Adaptive Frequency Switch, Adaptive Frequency Hopping, and Timeslot hopping. It can co-exist with other standards that are compliant with IEEE 802.15.4.

3.4 ZigBee

ZigBee [16, 19] is an open standard designed with the aim to provide an intelligent solution that meets wireless devices requirements in terms of cost, energy consumption, reliability, etc. ZigBee utilize the upper layer of the IEEE 802.15.4. ZigBee network supports mesh, star and tree topologies and are categorized into three device types, each has a specific role [20].

The main function of the Coordinator in all ZigBee networks is to set up the network, therefore it must be aware of all the constituent nodes and should be able to store information and manage communication and security keys of the network. Routers are the intermediaries, which facilitate information flow between devices. The endpoint devices are required to perform a limited function of interacting with their parent nodes.

3.5 The Highway Addressable Remote Transducer (HART)

The Highway Addressable Remote Transducer (HART) [16] is an industrial standard developed on the basis of IEEE 802.15.4 specification using up to 15 different

channels and a frequency of 2.4 GHz [21]. WirelessHART was developed to support power efficiency, reliability and integrated security. It uses Time Division Multiple Access (TDMA) with Frequency Hopping Spread Spectrum (FHSS) to access different channels at different time slots, while blocking channels which experience interference with the signal by creating a Blacklisting. These mechanisms will result in less interference with the existing wireless systems and reduce noise impact [18, 22, 23].

A typical WirelessHART network that the network manager can support are mesh, star and combination of both topologies and it consists of four components that are essential to provide a fully functional network [20, 22–24].

3.6 ISA100.11a

This standard designed on the basis of the IEEE 802.15.4 with the aim to provide reliability to the wireless communications infrastructure for monitoring and control different types of applications [25–27]. ISA100.11a like WirelessHART supports channel hopping and channel blacklisting to reduce interference and allows a device to coexist with other RF devices in the same band. It also benefits from TDMA mechanism that enables a device-to-device communication without delay in accessing an RF medium using the allocated time slot of 10 ms [22, 23]. At the network layer the standard facilitates the use of 6LoWPAN that is needed to handle IPV6 traffic and thus gives users the opportunity to connect to the internet. The ISA100.11a standard supports symmetric AES 128bit encryption.

An ISA100.11a network supports star, mesh and a combination of both and facilitates co-existence with wirelessHART [23] and it consists of different components that are essential to provide a fully functional network detailed in [22].

The authors in [28] have conducted study comparing WirelessHART and ZigBee in terms of some quality of services metrics such as robustness, coexistence and security. The outcome of the study showed that WirelessHART is a better choice for industrial applications compared to ZigBee in some quality of services terms. The authors in [17] have performed a comparative study on the WSN standards (wire-lessHART, ISA100.11a and WIA-PA) [13].

3.7 Discussion

Based on the above review on WSN standards we can conclude that wirelessHART and ISA100.11a both are suitable for industrial and larg scale deployment, while ZigBee technology is suitable for general applications due to the limitation on radio range of 100 m. WIA-PA is relatively new and it is still not widespread deployed.

WirelessHART and ISA100.1 both standards define the radio on the basis of the IEEE 802.15.4 standard for low-rate wireless personal area networks (L-R WPAN).

Table 1 A comparison between WSN standards [2, 25, 29]

Features	ZigBee	WirelessHART	ISA100.1
Underlying Standard	IEEE802.15.4	IEEE802.15.4	IEEE802.15.4
Security	High Symmetric	High Symmetric	High Symmetric/ Asymmetric
Scalability	Yes	Yes	Yes
Power consumption	High	Low	Low
Latency	Low	High	High
Co-Existence	NO	Yes	Yes
Application	Commercial	Industrial	Industrial
Bandwidth	20–250 Kbps		
Transmission Range	2.4 GHz and	2.4 GHz	2.4 GHz Free Band
Network Size	65,000	Up to 250 nodes on a network	Above 250 nodes per network
Topology	Mesh	Mesh and Star	Mesh and Star
Radio Channel	CSMA-CD	TDMA	TDMA/CSMA-CD
Network Routing Strategy	AODV/Tree Routing	Graph/Source/ Superframe Routing	Addressing; Routing

WirelessHART also uses unmodified IEEE802.15.4-2006 MAC, while ISA100.11a uses a modified, flexible IEEE802.15.4-2006 MAC. Both use the same mechanisms for establishing the wireless network and exchange data between network devices. Both standards deploy energy saving mechanisms that offer batteries longer life-time. Both standards survive the interference caused by the presence of other network system by combining channel-hopping and direct-sequence, spread spectrum (DSSS). The TDMA-based mesh topology used by WirelessHART and ISA100.1 provide a centralized infrastructure, that is required for large scale deployment. This makes both standards preferred ZigBee with its tree topology. WirelessHART and ISA100.11a differ from each other in their aims and objectives they designed to achieve.

Table 1 summarizes the various quality of services requirements of the previously discussed standards.

4 Requirements of Wireless Sensor Networks

There are certain requirements that are essential for a successful deployment of WSNs in the oil, gas and resources industries. In this section, the major requirements are dis-cussed. In Fig. 3, we propose a taxonomy of the requirements of WSN in the oil and gas industry.

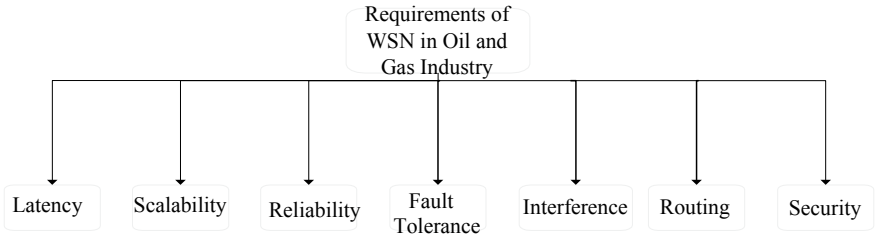


Fig. 3 Taxonomy of requirements of WSN in the oil and gas industry

4.1 Latency and Scalability

Latency is a measure of delay in time [30] that is requires a data packet to succefully reach the receiving node and acknowledged, it is measured in ms. The probability of a succefull transmission depends largely on the quality of transmission media, which influenced by the signal-to-noise ratio in the RF domain. The number of hops from source to destination can also affect the latency [29]. **Scalability** is the ability of the network to cope with the growth in number of the sensor nodes which vary from a few to several thousands. Different requirements of applications in oil and gas industry require protocols and standards that are scalable and able to adapt new functionalities such as adding and removing numerous sensor nodes without impacting the quality of service (QoS) performance metrics [18, 29].

4.2 Fault Tolerance and Reliability

Fault tolerant is described as the capability of the network to operate in a reliable fashion and adapt any topological changes that might occur as a result of node failures and link failures. Many factors [22] can affect the performance of a sensor node and cause frequent failures, such as energy depletion, hash physical location and other environmental effect [31]. Reliability is the measure of the amount of data that is delivered from source and accurately received by the destination with minimum possible packet loss [29, 32, 33]. Reliability can be defined by the application. For instance: reliability for data transfer is the integrity of data and all information are accurately received by the receiver; reliability for audio or video, reliability concerns more of tolerable distortion at the application layer. Usually the reliability in each layer of the communication stack [18] has various requirements and definition of error rates, error burst, delay, error concealment techniques, etc.

4.3 Co-Existence and Interference

WSNs transmit on low-power signals are extremely sensitive to noise, which makes the radio signal subject to interference multi-path distortion with other wireless networks and communication systems operating under the same radio-frequency (RF) medium. WSN standards should be able to function efficiently and co-exist in the occurrence of interference [18, 34, 35]. The authors [35] have surveyed the impact of interference on WSN working under the 2.4 GHz frequency and the existing approaches to tackle it.

4.4 Routing

An essential design criteria is the capability to deliver packets from source to destination consuming the least amount of energy possible. There are various constraints that must be taken into account when designing a routing protocol for WSNs, such as node deployments, energy, scalability and other quality of services necessary metrics for reliable data delivery. It is therefore important to overcome these challenges before efficient communication and reliable monitoring system for oil and gas industrial applications can be achieved in WSNs [36].

4.5 Security

One of the major benefits of WSNs is the ability to place sensor nodes in an environment without any supervision. This can provide security drawbacks to the network and backend system if the sensor nodes are located in harsh environments or in an unsecured manner while being readily accessible to people. Another limitation that makes WSN more vulnerable to eavesdropping and security breaches than its wired counterpart is the wireless transmission of data. Therefore there is a need for security system that protect access and privacy [22].

4.6 Power Consumption

The Oil and Gas Industry urges longer battery lifetimes that lasts over years for wireless sensors as a long term solution for the enormous effort required to maintain the network when replacing depleted batteries. In WSNs rechargeable batteries the main source of power for sensor devices, a regular replacement of the batteries is complex due to the number of several thousand sensors per plant, therefore nodes need to adapt an energy-aware mechanism in order to rationalize the consumption

of energy and thus ensure a longer lifetime of the network. Power consumption also plays an important for the routing of packets between the sensor nodes and must be considered by the routing protocol design criteria strike a balance between energy savings and functionality [25].

5 Existing Solutions

5.1 Pipeline Monitoring

In [37], the authors proposed a magnetic-induction based pipeline monitoring system (MISE-PIPE) that is intended to provide real time leak detection in both underground and above ground pipeline. A performance evaluation test showed precise results in detecting and localizing leakages. In [38], a framework that aimed to provide cost-effectiveness, security and reliability in monitoring pipeline structures is proposed. However the work does not provide neither technical design details nor performance tests. A crude oil leak detection and location monitoring system has been proposed by the authors in [39]. A wired PLC-5 programmable controller and wireless 1785-KE card were used to provide connectivity to the internet. It has proven its stability and reliability by detecting a leak that is larger than 2% of the total flow in 180 seconds, and accurate position that is less than 2% of the length of the pipeline. The authors in [40] introduced a heterogeneous network for underwater monitoring of oil and gas production to detect oil and gas leakages and to improve the flow and well production. A combination of ultrasonic and optical networks is used in additions to wires, to provide redundancy. Results showed that, connecting Tmote nodes with Micro-Electro-Mechanical Sensors (MEMS) to other Tmotes in the network using frequency (RF) communication or ultrasound or light connection improve the device flexibility.

A system that provides an early detection and warning ahead in time of failures and unexpected events has been proposed by the authors in [41]. The aims of the system are to meet the requirement of a real time monitoring in terms of the time delay between discovering a failure, examining the cause and preventing damage. The results show that 80% of messages received in less than 1ms, where 90% have an average of 10ms latency. The authors in [42] have proposed an online monitoring and inventory management system (TOTE INVENTORY MANAGEMENT) based on Radio Frequency Identification (RFID) and WS that helps monitor production activities in oil and gas plants. The results show that the system will improve the operation, however the proposed system might face several challenges such as power consumption and interference with other in plant existing system.

The authors in [43] have tackled the problem of the safety in gas fields by introducing a multi-sensor gas monitoring platform that detects possible gas leakage in reliable manners and provides an accurate positioning of the source. The authors in [32] have addressed the moisture problem related to LPG, LNG pipeline in oil

and gas production plant by introducing a monitoring system that is equipped with RFIDs sensor to detect the level of the received signal energy in a certain amount of time and thus the existence of moisture in the pipeline is determined.

5.2 Fault Tolerance, Co-Existence and Interference

The authors in [44] proposed a WSN application to monitor shipboard PdM aboard an operating oil tanker. They have conducted an experiment using a mesh topology with three Intel Mote clusters and three Mica2 Mote clusters were used to investigate the fault tolerance problem. A fault tolerance solution, called FTSHM (fault-tolerance in SHM is proposed in [45] the authors aim is to predict the node or link failure and repair them before they occur. The authors in [46] have proposed a network infrastructure that combines the features of both wired and wireless sensor networks. The wired network devices act as primary connection and the wireless network sensors providing a backup in the presence of network failures. They tackled the fault tolerance and power constraint by designing a network system in which the connection between nodes is performed through wireless transceiver and wired links, while the wireless transceiver are turned off for future backup activities the communication occur using the wired links. The power is provided through wired Devices. The authors in [47, 48] carried out a research about the technical possibility of implementing WSN in oil and gas industry environments. For they used WSN standards (WirelessHart, ISAl00.11, ZigBee) as solutions for applications monitoring. The results showed that all three studied standards function efficiently in the presence of other network systems and are able to survive the interference and noise issues.

5.3 Routing, Energy Consumption and Security

An enhanced version of the original AODV routing protocol has been proposed by the authors in [36]. In the M-AODV the HELLO and ERR messages exchanged by the neighbor's device in the original AODV to announce their presence and update the routing table are eliminated and thus flooding the network with unnecessary broadcasts is avoided. The results obtained from the comparative study between M-AODV against AODV and HERA protocol showed the A-AODV outperformed both protocols interms of delay, but has similar performance to HERA in terms of stability and collision. It is also found that M-AODV has a higher transmission that reaches 99% in 3 minutes time interval 4 times faster as the original protocol. A study of Sensor life time maximization with respect to sensor placement along a pipeline under different power model, ideal power model and Tmote power model has been investigated by the authors in [49]. For their study, they distributed a set of Tmote Sky sensors with low power MCU MSP430, TI CC1101 transceiver

chip. The sensors were distributed uniformly with equal distance. It is found that an increasing in the number of nodes under the equal distance scheme will result in decreasing the lifetime. The authors in [50] have tackled the security issues faced when deploying WSNs to control and monitor the safety of crude oil pipelines in industrial automation, specifically the Niger Delta region of Nigeria. They proposed a security solution that could protect oil and gas facilities from Vandalism by enabling a reliable and fast detection and reporting of possible security breaches. They used Cyber Physical Systems (CPS). Table 2 provides a summary of the existing solutions that tackle the challenges and enhance the deployments of WSNs in oil and gas industry.

Table 2 Summary of existing solutions

Study	Technical requirements	Technology used	Types of Motes	Network type
[2]	Energy consumption and Reliability	WirelessHart	N/A	Mesh and a SensiNet network
[5]	Reliability, power consumptions and fault tolerance	Multi-hop routing algorithm	N/A	Linear infrastructure
[6]	Gas leaks in pipeline localization	ZigBee protocol	Temperature sensor	Linear infrastructure
[7]	Gas leaks, corrosion, H_2S	CDMA/GPRS	Temperature, pressure, motor current and voltage	Redundant multi-hop topology
[30]	Reliability	ATEX version of DUST wireless networks	N/A	N/A
[36]	Stability and reliability	M-AODV (modified AODV)	N/A	Mesh topology
[51]	Gas pipeline pressure, temperature	Optical fiber	Pressure and temperature	Distributed
[37]	Gas pipeline leakage	Pressure, acoustic and soil property sensors	N/A	Magnetic induction-based wireless sensor networks
[38]	Pipeline monitoring scalability	N/A	MicaZ mote, RF Chipcon and CC2420 transceiver	N/A
[39]	Stability and reliability in Gas leakage detection	MODEM, GPS and wireless 1785-KE card	Pressure, flow and temperature	N/A

(continued)

Table 2 (continued)

Study	Technical requirements	Technology used	Types of Motes	Network type
[52]	Leakage detection reliability	Wave	N/A	N/A
[40]	Redundancy in underwater gas Leakage monitoring	Ultrasonic and optical networks	Moteiv Tmote Sky Type) with TinyOS	Heterogeneous
[41]	Time delay in Gas Leakage detection	SOAP and MQTT	N/A	N/A
[42]	Monitoring and management inventory system	WirelessHART and RFID	N/A	Mesh topology
[43]	Safety in Oil and Gas Field	GPS\FRID, STM32F103 and μC/OS-II	Gas sensors	N/A
[32]	Moisture in LPG, LNG pipeline	Pico-RFID and Agilent's PSA E4440A	N/A	N/A
[53]	Latency and Data rates	N/A	Tmote sky devices, MIMO access AGN1200 and TS-3300 board	Heterogeneous
[44]	Fault-tolerance	N/A	Intel Mote and Mica2 Mote	Mesh topology
[45]	Reliability, energy consumption and interference	backup sensor placement (BSP)	N/A	N/A
[46]	Fault-Tolerance	N/A	N/A	Wired/Wireless sensor network
[47]	Interference	WirelessHart, ISAl00.11, ZigBee	N/A	N/A
[48]	Interference	WirelessHart, ISAl00.11, ZigBee	N/A	N/A
[49]	Energy consumption	N/A	Tmote Sky sensors with low power MCU MSP430, TI CC1101	Uniformly distributed nodes
[50]	Security	CPS, Wifi	N/A	N/A

6 Conclusion

For Oil and Gas Industry WSNs offer a large amount of applications that are useful for the production of Performance Optimization (monitor pipelines, gas detection, corrosion, H_2S, equipment health status, and real-time reservoir and process control, safety, maintenance. The technical capability of wireless sensor networks (WSN) is still under ongoing exploration in science, as its true potential has not been fully exploited. This paper aimed to provide WSN designers, oil and gas companies and researcher a critical review at the provision of WSN in the oil and gas industry taking into account the specific requirements of the applications and challenges. Furthermore, a particular attention is given to the existing architectures of efficient mechanisms that fulfil the requirements and overcome the challenges that arise when deploying a WSN in Oil and Gas industry.

Acknowledgements This work was supported by Edinburgh Napier University and the King Abdulaziz City for Science and Technology (KACST) within the RsIoT project number: AT-35-98

References

1. Akhondi MR, Talevski A, Carlsen S, Petersen S (2010) Applications of wireless sensor networks in the oil, gas and resources industries. In: 24th IEEE international conference on advanced information networking and applications, Perth
2. Peterson S, Doyle S, Vatland C, Aasland T, Tman T, Sjong D (2007) Requirements, drivers and analysis of wireless sensor network solutions for the oil & gas industry. In: The IEEE conference on emerging technologies and factory automation, Patras
3. Talevski A, Carlsen S, Petersen S (2009) Research challenges in applying intelligent wireless sensors in the oil, gas and resources industries. In: 7th IEEE international conference on industrial informatics, Perth
4. Gutierrez JA (2005) IEEE STD. 802.15.4—Enabling pervasive wireless sensor networks. In: EATON corporation
5. Jawhar I, Mohamed N, Mohamed M, Aziz J (2008) A routing protocol and addressing scheme for oil, gas, and water pipeline monitoring using wireless sensor networks. In: 5th IFIP international conference on wireless and optical communication networks
6. Husnain S, Salman A, Sidra R, Qaisar S, Emad F (2014) Reliable monitoring of oil and gas pipelines using wireless sensor network (WSN)—REMONG. In: 9th international conference system of systems engineering (SOSE)
7. Akhondi MR, Talevski A, Carlsen S, Petersen S (2010) The role of wireless sensor networks (WSNs) in industrial oil and gas condition monitoring. In: 4th IEEE international conference, the digital ecosystems and technologies (DEST), April 2010
8. Rault T, Bouabdallah A, Challal Y (2014) Energy efficiency in wireless sensor networks: a top-down survey. In: Computer networks, Elsevier
9. Azubogu AC, Idigo VE, Nnebe SU, Oguejiofor Obinna S, Simon E (2013) Wireless sensor networks for long distance pipeline monitoring. Int J Electric Comput Energetic Electron Commun Eng
10. Chao X, Waltenegus D, Lin G (2008) Energy model for H2S monitoring wireless sensor network. In: 11th IEEE international conference on computational science and engineering. https://doi.org/10.1109/CSE.2008.33

11. Lennvall T, Svensson S, Hekland F (2008) A comparison of WirelessHART and Zigbee for industrial applications. In: IEEE international conference workshop factory communication system
12. Available: www.bluetooth.com/Pages/low-energy.aspx (2017) [Online]
13. Alhmiedat T (2015) A survey on environmental monitoring systems using wireless sensor networks. J Netw 10(11)
14. Available: http://www.ieee802.org/15/pub/TG4.html (2017) [Online]
15. Chaari L, Kamoun L (2011) Performance analysis of IEEE 802.15.4/Zigbee standard under real time constraints. Int J Comput Netw Commun (IJCNC) 3(5):9
16. Wang JJQ (2016) Comparative examination on architecture and protocol of industrial wireless sensor network standards. IEEE Commun Surv Tutor PP(99):1
17. Liang W, Zhang X, Xiao Y, Wang F, Zeng P, Yu H (2011) Survey and experiments of WIA-PA specification of industrial wireless network. Wirel Commun Mob Comput 11 (8):1197–1212
18. Ajith Kumar SA, Øvsthus K, Kristensen LM (2014) An industrial perspective on wireless sensor networks—a survey of requirements, protocols and challenges. IEEE Commun Surv Tutor 16(3)1391–1412
19. Available: http://www.zigbee.org (2017) [Online]
20. Vandana J, Verma AK (2015) The underlying technologies in WSNs: ZigBee vs. Wireless HART. In: 12th international conference on fuzzy systems and knowledge discovery (FSKD)
21. Available: http://www.hartcomm.org (2017) [Online]
22. Radmand P, Talevski A, Petersen S, Carlsen S (2010) Comparison of industrial WSN standards. In: 4th IEEE international conference of the digital ecosystems and technologies (DEST)
23. Labeau F, Agarwal A, Agba B (2015) Comparative study of wireless sensor network standards for application in electrical substations. In: Computing, communication and security (ICCCS), international conference
24. Remke A, Wu X (2013) WirelessHART modeling and performance evaluation. In: 43rd annual IEEE/IFIP international conference on dependable systems and networks (DSN)
25. Rault T, Bouabdallah A, Challal Y (2014) Energy efficiency in wireless sensor networks: a top-down survey. Comput Netw 67:104–122 (Elsevier)
26. Available: http://www.isa.orgISA100 (2017) [Online]
27. Werb J (2014) ISA100 wireless applications, technology, and systems A Tutorial white paper
28. Lennvall T, Svensson S, Hekland F (2008) A comparison of WirelessHART and Zigbee for industrial applications. In: IEEE international conference workshop factory communication system, pp 85–88
29. Nechibvute A, Mudzingwa C (2013) Wireless sensor networks for SCADA and industrial control systems. Int J Eng Technol 3(12):12
30. Carlsen S, StatoilHydro N, Skavhaug A, Petersen S, Doyle P (2008) Using wireless sensor networks to enable increased oil recovery. In: IEEE international conference on emerging technologies and factory automation
31. Ma R, Xing L, Michel HE (2006) Fault-intrusion tolerant techniques in wireless sensor networks. In: Dependable, autonomic and secure computing, 2nd IEEE international symposium
32. Zhu L, Zou B, Zhang H, Wang Z, Jiang M (2015) Design of multi-sensor wireless monitoring system and its application in natural gas purification plant. In: IEEE international conference on mechatronics and automation (ICMA)
33. Deb B, Bhatnagar S, Nath B (2003) ReInForM: reliable information forwarding using multiple paths in sensor networks. In: 28th Annual IEEE international conference on local computer network
34. Werb J, Berry V, Lamb S, Sexton D, Lapinski M, Newman M (2005) Improved quality of service in IEEE 802.15. 4 mesh networks. In: International workshop on wireless and industrial automation

35. Willig A, Matheus K, Wolisz A (2005) Wireless technology in industrial networks. In: IEEE
36. Ivanovitch S, Affonso LG, Francisco V (2010) A new AODV-based routing protocol adequate for monitoring applications in oil & gas production environments. In: 8th IEEE international workshop of the factory communication systems (WFCS)
37. Zhi S, Pu W, Mehmet CV, Mznah AR, Abdullah MAD, Ian FA (2011) MISE-PIPE: magnetic induction-based wireless sensor networks for underground pipeline monitoring. Ad Hoc Netw 9(3):218–227
38. Jawhar I, Mohamed N, Shuaib K (2007) A framework for pipeline infrastructure monitoring using wireless sensor networks. In: Wireless telecommunications symposium
39. Pakzad S, Fenves G, Kim S, Culler D (2008) Design and implementation of scalable wireless sensor network for structural monitoring. In: New sensors, instrumentation, and signal interpretation
40. Covas D, Ramos H, de Almeida A (2012) Standing wave difference method for leak detection in pipeline systems. J Hydraul Eng 13:1106–1116
41. Dalbro M, Eikeland E, in't Veld A, Gjessing S, Lande T, Riis H, Sørasen O (2008) Wireless sensor networks for off-shore oil and gas installations. In: Second international conference on sensor technologies and applications IEEE (SENSORCOMM)
42. Hill M, Campbell M, Chang Y, Iyengar V (2008) Event detection in sensor networks for modern oil fields. In: Second international conference on Distributed event-based systems, ACM
43. Vellingiri S, Ray A, Kande M (2013) Wireless infrastructure for oil and gas inventory management. In: 39th annual conference of the IEEE industrial electronics society, IECON
44. Johnstone I, Nicholson I, Shehzad B, Slipp J (2007) Experiences from a wireless sensor network deployment in a petroleum environment. In: International conference on Wireless communications and mobile computing (IWCMC), New York, USA
45. Krishnamurthy L, Adler R, Buonadonna P, Chhabra J, Flanigan M, Kushalnagar N, Nachman L, Yarvis M (2005) DesDesign and deployment of industrial sensor networks: experiences from a semiconductor plant and the North Sea. In: 3rd international conference on the embedded networked sensor systems (SenSys), ACM, New York
46. Bhuiyan MZA, Wang G, Cao J, Wu J (2015) Deploying wireless sensor networks with fault —tolerance for structural health monitoring. IEEE Trans Comput 64(2):382–395
47. Mohamed N, Jawhar I (2008) A fault tolerant wired/wireless sensor network architecture for monitoring pipeline infrastructures. In: Second international conference of the sensor technologies and applications (SENSORCOMM)
48. Skavhaug A, Myhre B, Sjong D, Mikkelsen E, Hendrik J, Doyle P, Sansom M, Carlsen S, Petersen S (2008) A survey of wireless technology for the oil and gas industry. In: Intelligent energy, RAI. Society of Petroleum Engineer, Amsterdam
49. Guo Y, Kong F, Zhu D, Şaman Tosun A, Deng Q (2010) Sensor placement for lifetime maximization in monitoring oil pipelines. In: 1st ACM/IEEE international conference on cyber-physical systems (ACM)
50. Obodoeze C, Ozioko F, Mba C, Okoye F, Asogwa S (2013) Wireless sensor networks (WSNs) in industrial automation: case study of Nigeria oil and gas industry. Int J Eng Res Technol (IJERT) 2(3):1–7
51. Alan D (2000) Optical fiber sensors for permanent downwell monitoring applications in the oil and gas industry. IEICE Trans Electron E83-C(3):400–404
52. Li W, Zhu Y (2006) Analysis on leakage detection and location techniques for long transmission Pipeline. IEEE Comput Soc 7(3)
53. Nasir A, Soong BH, Qaraqe KA (2013) RFID in-pipe moisture sensing system for oil and gas quality monitoring in Qatar. In: 19th IEEE international conference on networks (ICON)

Applying Human Behaviour Recognition in Cloud Authentication Method—A Review

Mohanaad Shakir

Abstract Authentication methods in public cloud computing play considerably in preventing unauthorized access to data saved in the public cloud. Based on the National Institute of Standards and Technology (NIST), cloud computing type public is a weak confidentiality and integrity sureties. Insufficient security controls in the cloud provider's platform could negatively affect the system's confidentiality and privacy. The aims of reviewing related works in this paper are collecting strong points and identifying the existing works' weaknesses on current authentication methods in public cloud computing from the behavior authentication side. The investigation shows that there are many Multifactor authentication methods for obtaining authentication in public cloud computing. The top three methods are password with a smart card, Mobile, and biometric. Many researchers have suggested an authentication model in public cloud computing using these methods. However, all these methods have suffered from many drawbacks and threats. This paper finds that the accuracy of the current Multifactor authentication is deficient in dealing with stolen password attacks in public cloud computing. In a related context, the best method to improve authentication accuracy is to obtain intelligent authentication operations.

Keywords Authentication · Cloud authentication · Behaviour recognition · Cloud security · Intelligent security · Human behaviour

1 Introduction

Based on (NIST), cloud computing is defined as "a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources that can be rapidly provisioned and released with minimal management effort or

M. Shakir (✉)
College of Business (CoB), University of Buraimi (UOB), Buraimi, Oman
e-mail: mohanaad.t@uob.edu.om

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_45

service provider interaction" [1]. It is commonly broken down into three deployments: public, private, and hybrid [2].

Public cloud computing is weak when facing a stolen password attack. For example, using a stolen password of remote access through an authentication layer could allow an unauthorized user to gain unauthorized access [3]. Authentication methods play a considerable role in preventing unauthorized access to data saved in the public cloud [3]. It is divided into many types such as; Single Sign-On, Mobile Trusted, Username and Password Authentication, Public Key Infrastructure, Biometric Authentication, implicit authentication, and Multifactor Authentication (MFA). Multifactor Authentication (MFA), which has many sub-categories, would be difficult to crack in a similar situation. MFA has many classical models, such as a password with a smart card, pin, SMS, fingerprint, face recognition, and behavior recognition [4, 5].

In MFA, the behavior recognition technique is considered cheap and easy to use. Many researchers recommend the current approaches of password-based user authentication research to be improved by incorporating behavior recognition in human cognitive factors both in design and run-time [6–16]. Besides, several current studies have focused on the authorized behavior users that interact inside the system with web-based software, such as moving the mouse and clicking on the keyboard [6–15]. However, this process has accentuated the problem of unauthorized users having little time to deal with data saved in the public cloud until it is diagnosed. Moreover, previous studies of behavior recognition to authenticate an authorized user in public cloud computing have not dealt with an authorized user who interacts with a password before accessing the system. However, the literature has emphasized the importance of moving from traditional security processes to intelligent security processes [17, 18].

Intelligent security is defined as information related to protecting an organization from external and internal threats through apply intelligent methods. The processes, policies, and resources are designed to gather and analyze that information as intelligent protection [19, 20]. There are many intelligent mechanism applications in the authentication process, such as Neural network in Leonard [6], Leaning Algorithm in Shi et al. [21], and agent in Mostafa et al. [22], Mohanaad et al. [23], Mohanaad [24]. Hajivali and Zhang [22, 25] have recommended for the intelligent actions of user authentication to be applied on public cloud computing. The framework discusses authentication in public cloud computing (PCC) concepts such as authentication in public cloud computing, authentication definitions, services, characteristics, authentication approaches features, and the limitation of current methods. To know the research areas currently studied by researchers, reviews on research areas in authentication in public cloud computing are also conducted and documented. Besides, The majority of recent studies have ignored human behavior as a factor in authentication. They also have neglected the fact that intelligence in password-based authentication is feeble. This paper aims to give the most research overview in authentication in PCC and the current research areas that have focused on many researchers, such as Multifactor authentication in PCC, human behavior recognition in authentication, and password-based authentication in PCC.

2 Research Method

An overview of the overall issues discussed in this paper can be viewed using the conceptual framework presented in Fig. 1. The framework discusses authentication in public cloud computing (PCC) concepts such as authentication in public cloud computing, authentication definitions, services, characteristics, authentication approaches features and the limitation of current approaches. To know the research areas currently studied by researchers, reviews on research areas in authentication in public cloud computing are also conducted and documented.

3 Authentication Methods in Public Cloud

Public cloud computing authentication involves validating the user's identity, ensuring that he is authorized to gain access to public cloud computing [24, 54]. Authentication is a vital feature of security compliance methods in public computing. It safeguards users against current security and privacy threats by preventing unauthorized access to public cloud user data [26, 27]. According to the Correlation Matrix of Latent Variables, the author performs an analysis to assess public cloud computing CHALLENGE. According to the findings of this report, the diagnosis of the approved user for cloud access is the most worrying problem among the various types of security risks in cloud computing [28]. In cloud computing, the authentication layer is used to verify an authenticated user's identity and allow him access to data stored in the cloud. The seven forms of Authentication in public cloud

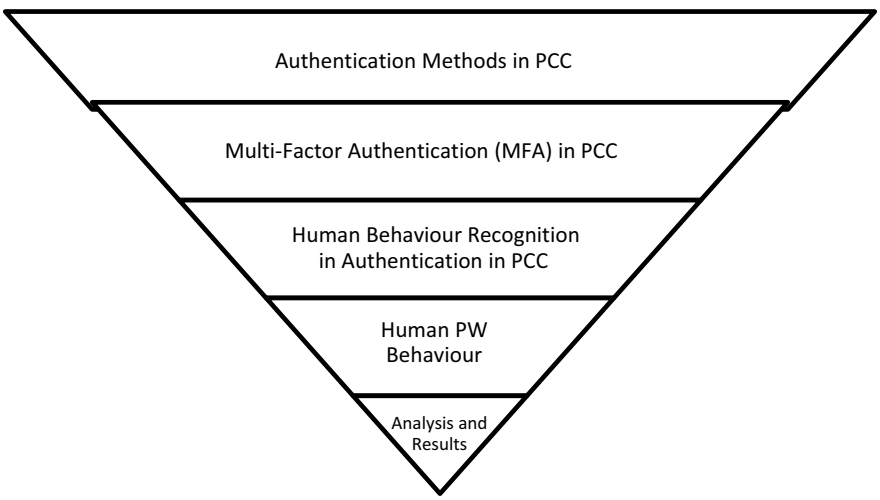


Fig. 1 Conceptual framework

computing (PCC) are Username and Password Authentication, Multifactor. Authentication, Mobile.Trusted, Single.Sign-On, Public. Key Infrastructure, Biometric.Authentication and Implicit. Authentication as shown in Table 1. Despite the growing number of novel authenticate users, the password-based authentication method remains the most popular choice [5, 24, 29].

Multifactor authentication (MFA) method that aims to enhance the security of different websites has gained popularity [30]. Its two sub-categories would be more difficult to crack in than Password-based authentication alone. In addition, its security is more guaranteed as it also requires another factor, instead of just validating the username and password pair [31]. Many researchers such As Ravi Kumar [32], Matthew et al. [33], Muhammad et al. [34], Sumitra et al. [35], Alok Tripathi et al. [36], and Nalini et al. [37] state: there are several reasons why Multifactor authentication (MFA) is a successful tool for reducing cloud authentication attacks.

4 Multi-Factor Authentication (MFA)

In ensuring data is more secure in the public cloud computing, there needs to be a combination of authentication techniques employed. This method would promise more security because it requires another factor, e.g. biometric authentication, as opposed to just validating the username and password pair. It proposes to be one of the stronger authentication techniques. Authentication mechanisms currently use include:

1. Passwords.
2. Smart cards.
3. Digital certificates.
4. Kerberos.
5. Biometrics.

The following are the three traditional ways of authentication:

1. Something the user is aware of, such as a password or a pin.
2. Something the user possesses, such as a smart card.
3. Something about the user, such as an iris scan or a fingerprint.

The expectation of authenticity increases exponentially as additional variables are used in the verification process. For the cloud computing environment, a multifactor biometric authentication method that includes fingerprint and palm vein is proposed [4]. This is done to manage biometric data securely, with fingerprint data stored in a central database of cloud security services and palm vein biometric data stored in multi-component smart cards. However, there are limitations when facing stolen password attack. In addition, extra hardware has been needed to apply would make it costly if an entire organization is to use this feature to authenticate users, except behaviour recognition.

Table 1 Authentication types

Authentication type	Description
Username and password authentication	Confidentiality and privacy can be maintained at some level in this authentication technique. For the information to be accessed in the CSP, the user needs to enter the username and password. As it is difficult to determine whether the request is from the authorized user, this technique does not seem to provide higher and reliable security
Multifactor authentication (MFA)	Access is given only after the user provides two or more pieces of evidence (or factors) for a legitimate authentication mechanism: information (something the user and only the user knows), ownership (something the user and only the user has), and intrinsic (something the user and only the user has) (something the user and only the user is)
Mobile trusted	The Trusted Computing Group (TCG) implements a collection of conditions that uses a hardware root-of-trust (Mobile Trusted Module (MTM) and Trusted Platform Module (TPM)) to store, calculate, and report software and hardware integrity (TPM). If TPM is designed for PCs, MTM is designed for mobile devices [60]. The credibility and stability of a mobile platform are guaranteed with MTM [61]
Single sign on (SSO)	Single Sign On (SSO) is a method of gaining access to multiple independent software systems in a way that a user is able to have access to all the systems without being required to re-login in each application [62]
Public key infrastructure (PKI)	Traditional authentication methods, such as RSA, depend on the secret key and are designed to support the use of traditional asymmetric cryptographic algorithms. A private key is used to verify a user's identity [33]
Biometric authentication	The process of validation is considered complete if a user is indeed who he claims to be. In biometric authentication, there are three important factors of information security: identification, authentication and non-repudiation. This authentication technique is based on recognition of an individual's behavioural and physiological features. Its ability to provide the biological proof of what we are and what we know makes biometric authentication a strong authentication technique [63]
Implicit authentication	This method authenticates a user by monitoring his behavior, and it is appropriate for mobile devices since they are capable of collecting a large amount of user data, such as location, motion, contact, and application use. In order to evaluate a suitable service for user and personal profile information in the mobile cloud world, a range of profiling techniques have been investigated [64–66]

5 Stolen Password Attack

An attacker explicitly obtains a user's login credentials, or more precisely, his or her password, in the most simple threat model that can occur to a user's accounts [38]. Many well-known online and offline assaults, such as client-side malware, phishing with a spoofed site, and eavesdropping on password transmission, can be used to accomplish this. When used, these attack tactics will compromise a user's account for a specific site, as well as any personal information associated with it. According to Mark Zuckerberg [39], Facebook was caught up in an estimated 272 million stolen password attacks in only one month. However, in recent years, many protections against this threat have grown. In addition to a password, this involves using multiple-factor authentication, such as SMS, Smartcard, and Biometric [40]. Nonetheless, limitations to these models exist because of the additional hardware needed to implement this technique, which can become costly if an entire organization uses this feature to authenticate users. In addition, If a user's authentication factor and password are both compromised simultaneously, an attacker would be able to impersonate the user [41]. In a related context, many researchers are recommended to apply behavior recognition as an authenticating factor with a password during the authentication process [6–15, 20, 23]. Thus, in this paper, I select behavior recognition with the password to improve the accuracy in authenticating an authorized user when facing a stolen password attack [30–42].

6 Human Behaviour Recognition in Authentication

Several researchers have proposed a method of authentication that combines behavioral recognition with the human element. Several studies have used biometrics, such as cognitive fingerprints, eye scans, clothing color, and facial tracking, to authenticate users continuously [8, 43, 44]. However, to work effectively, many of these methods necessitate extra hardware and higher costs. Through studying how users communicate with the system, behavioral modeling tackles these limitations. Some of the methods used to monitor a user's actions continuously include evaluating mouse activity, assessing how users search for and select details, and determining the habitual typing rhythm of users [9, 10]. Although these approaches do not require special hardware, most of them require the installation of specialised monitoring software.

Most of the Multifactor authentication methods in public cloud computing have a high cost of implementation and deployment. Most of the current authentication approaches are weak when facing stolen password attack, which then leads to adverse effects on the accuracy of authenticating an authorized user [45–54]. Still, many researchers have suggested an authentication approach that uses behavioral recognition with human factor [6–15]. Most current studies neglect the presence of human behavior recognition in the authentication process as a factor in the

performance of authenticating an authorized user in public cloud computing [55]. In addition, learnability is highly weak in password-based authentication [55, 56]. Thus, in the following subsection, the related works in behavior recognition with the human in the authentication area are presented.

7 Behaviour Recognition in Authentication Related Work

Chow in [57] proposes an authentication method under the name TrustCube [58] according to implicit authentication [21] for mobile authentication (the name TrustCube is featured in both the initial and extended method). TrustCube, a policy-based cloud-based authentication solution that uses an open standard. It also makes use of a federated authentication scheme, similar to OpenID. The implicit authentication algorithms are not listed, but a top-level framework overview is given. Developed with implicit authentication, it utilizes mobile data like SMS messages, calling logs, location, and website access in the existing public cloud environment. However, it also has its limitations. Among them is, the public cloud constraints in input requirements will make it more difficult to use complicated passwords, and this, in turn, will lead to the use of short password and PINs. Consequently, this poses higher security risks such as stolen password attack. Niinuma et al. suggest a CUA framework that automatically registers a user's colour of clothing and face as soft biometric traits [8, 43]. This method can authenticate users irrespective of their status in front of the workstation. It uses information about users' colour of clothing as an enrolment template in addition to their facial information. Limitations in the study come from the additional hardware that needs to be implemented, as it can be costly should an entire organization decide to use this feature to authenticate users [6]. Altinok et al. suggest a continuous biometric authentic system that gives an estimation of authentication certainty at any given time, even when any biometric data is not present [11]. The study presents an initial approach for temporal integration depending on uncertain propagation over time to estimate channel output distribution from recent history, and classification with uncertainty. Some of the limitations of this study include the fact that the authentication uncertainty rises over time which will cause a decrease in system usability. This technique also requires hardware such as camera and scanner for fingerprints, which can become costly. Kang at el. present temporal integration of biometrics and behavioural features to authenticate users continuously [59]. Shen et al. employ mouse dynamics when carrying out continuous user authentication [9]. This technique is used to monitor behavioural features in mouse operations to recognise malicious users. Nevertheless, this emerging approach has some existing limitations. Behavioural variability is shaped by human or environmental factors. For instance, a user's behaviour will be modified significantly if he or she switches software environments or experiences biological or emotional change.

Leonard [6] suggested a web-based behavioural modelling for Continuous User Authentication (CUA). The technique can go together with web applications for reliable and secure authentication. Some challenges occur when modelling the behaviour of users that interact with web-based software. Other than that, an approach to address the need for web-based continuous user authentication is also presented. Statistical Language is applied to classify users according to document classification, information retrieval, machine translation and speech recognition. This study establishes a novel keyword abstraction process to identify user activity for each system user. It also develops new techniques to observe users' web-based applications behaviour using information that is already available to a system administrator. The main limitation of this technique is the time required for analyzing the authenticating user when s/he logins into the system by comparing current and previous user behaviour. This time gap enables an unauthorized user's illegal access to deal with data in the case of stolen password attacks. In addition, the monitoring of authenticity of the user interacting with a password, through behaviour, is neglected.

Most of the authentication methods that monitor an interacting user with password need extra hardware as an authentication tool in authentication process [8, 11, 43, 57]. However, these methods are too expensive to be used for authentication in public cloud computing [60]. On the other hand, the cost of authentication methods that monitors an interacting user with the system via applying extra software as an authentication tool during the authentication process is low [6, 9, 10, 13]. However, the main limitation in these methods is there is some time for the user to deal with data which is saved in the public cloud until the user's legitimacy is determined. Furthermore, in most of these methods, the monitoring of the user's behaviour when interacting with the password has not been investigated sufficiently. To avoid these problems, the literature has emphasized the importance of moving from traditional security processes to intelligent security processes [17] through adapting intelligent mechanisms that represent an authorized user's behaviour, which can automatically prevent an unauthorized user [56, 61]. There are many intelligent mechanism applications in authentication process such as Neural network in Leonard [6] , Leaning Algorithm in Shi et al. [21], and agent in Mostafa et al. [22]. Finally, many researchers such as Hajivali and Zhang [22, 25], have recommended for the learning process actions of user authentication on public cloud computing to be applied. Thus, in this paper I sheds light on intelligent authentication in the next section.

8 Intelligent Authentication

Many previous studies apply intelligent methods in the authentication process such as Mostafa et al. [22], Elaine et al. [21], and Vadim [62]. This study proposes a method named ACUA (Access Control and User Authentication) that utilises a cloud-based software-as-a-service application depending on one client-based agent

and four cloud-based agents (Authentication, Access control, Cryptography, and Cloud manager). In order to increase the rate of intelligence during cloud computing communications, authors apply agents. This study is conducted with the definite objective to establish a secure algorithm during processes, services and communications. In this method, the main limitation is diving data in several parts and joining them to the main part in the access control agent is time-consuming. Another limitation of this method is finding a suitable way to teach users to use the method without mistakes. Finally, this method does not fully represent the interaction among human factor with a password and web-based system.

Elaine et al. [21] state in this study the authors suggest implicit authentication; authenticating users based on behaviour patterns in cloud computing. This method consists of two major components. The first step learns about a user model through a learning algorithm based on past users' behaviour. In the second step, the user model characterises the user's behavioural patterns based on modelling independent features such as time elapsed since last good call, frequency of bad calls incidents per day and GPS coordinates. However, this algorithm is fairly robust to the pollution of a single feature. This algorithm has also introduced the learning algorithm principle of a user's behaviour as a new authentication factor. This method has proven the effectiveness of learning principle when it is embedded with authentication. As with many other methods, this method also suffers from many limitations and one example is the fact that it is suited for mobile devices rather than computers since it generates user behaviour patterns according to mobile parameters (time calls, bad calls, and location). In addition, this method has not dealt with the way of determining an illegitimate user during cases of stolen password and when the user tries to log in through the same time, device, and location. Finally, this method requires smartphones to generate users' behaviour pattern which leads to high cost when applied in any organisation.

Vadim [62] proposes a learning method in authentication to improve the accuracy of user authentication. The researcher in this study tries to apply the intelligent principles in the user authentication process. The study explains about employing JADE agents to share knowledge in a multi-agent system. This research defines three kinds of agent: Student, Teacher and Examiner. Student can receive messages from other agents and can choose to either adopt new knowledge or display the result of known action, but initially, it does not know anything. The second type of agent is "Teacher". Initially, it knows something. In this case, it knows the mathematical operation of adding on two elements of either integer or fraction. The teacher can also pass this knowledge during the communication act. The teacher is also able to identify any students before communication act happens because of the only agent with the ability to learn in Student. The third type of agent, which is the Examiner, knows the same terms that the Teacher does, but it searches for the Student agent specifically to inspect their knowledge by sending a special type of requests. The main limitations of this study the user needs to be monitored when interacting with password-based authentication.

9 Discussion and Conclusion

This paper covers public cloud computing authentication as a whole. In cloud computing, public deployment supports many applications from military to commercial applications. In addition, an authorized user can access the public cloud at any time anywhere by using any device through the authentication layer. Insufficient password security controls in public cloud computing could negatively affect the confidentiality and privacy of the system. For example, the use of the stolen password by remote access through the authentication layer could allow an unauthorized user to gain unauthorized access, modify, or destroy an organization's data [3]. Authentication is considered a millstone to allow an authorized user to deal with data saved in the public cloud [3, 63]. Authentication performance from an accuracy criterion is defined as the capability of the authentication system to correctly determine an authorized user [64, 65]. Password is still one of the most popular authentication methods of all [29], as it is considered the cheapest and the most common methods of computer authentication. Many researchers have reported password leaks as major issues in many well-known organizations, such as iCloud and Apple Inc. [66]. However, these password leaks are having a serious effect on the accuracy of user authentication when facing a stolen password attack [3, 67–69]. Multifactor authentication (MFA) model, such as a password with Smart card, SMS, and Biometric, is suggested by many researchers to avoid this problem [70]. The model of MFA is considered hard to break as compared to the use of password-based authentication alone. However, a set of deficiencies have been noted in the model of MFA such as high cost and stolen password attack which were discussed in details in above section. The gaps identified from the literature are used to investigate any potential areas in the accuracy of user authentication in a public cloud that requires further enhancements and modifications. The points below are gathered and summarized from the literature review:

1. Most current authentication approaches [6, 45, 48–50, 52–54, 71, 72] are weak when facing stolen password attack, which lead to negative effects on the accuracy of user authentication.
2. Most current studies have ignored the presence of human behavior as a matching or complementing factor in authentication [24]. They also have neglected the fact that intelligent in password-based authentication is highly weak [56, 73]. Therefore, when an unauthorized user is trying to log into a cloud through an authorized password, device, and network on the first attempt, the original user's authority will grant the unauthorized users' access to data saved in the cloud.
3. Many researchers have recommended intelligent mechanisms to represent human behavior in an authentication layer to improve the accuracy of user authentication [17, 56, 73–75].
4. Previous studies of human behavior recognition have not dealt with monitoring user when interacting with a password before access into the system [10, 14, 43, 76].

From reviewing the current state of this work, it was found that the accuracy of the current Multifactor authentication in public cloud computing is deficient in dealing with the stolen of password attacks [31, 35, 77–83]. In line with the above, the purpose of this paper is to highlight the gap related to the accuracy of user authentication. The weakness in most of the authentication process methods to face the stolen password attack affects negatively on the accuracy of user authentication [45–54]. This weakness is mainly due to its high cost of implementation and deployment. Besides, learnability from users behavior in authentication process is highly weak [55, 56]. All in all, in multifactor authentication within public cloud computing, the authentication process in password-based authentication process needs to move from traditional security process to intelligent security process [17] by developing an item. This item has an intelligent mechanism for representing an authorized user behavior that can prevent unauthorized user automatically [56, 61].

References

1. Kundra V (2011) Federal cloud computing strategy
2. Mell P, Grance T et al (2011) The NIST definition of cloud computing
3. Jansen W, Grance T (2011) Guidelines on security and privacy in public cloud computing. NIST Spec Publ 800–144:80
4. Ziyad S, Kannammal A (2014) A multifactor biometric authentication for the cloud. In: Computational intelligence, cyber security and computational models. Springer, pp 395–403
5. Eldow A, Shakir M, Talab MA, Muttar AKH, Tawafak RM (2019) Literature review of authentication layer for public cloud computing : a meta-analysis. J Theor Appl Inf Technol 97(12)
6. Leonard LC (2017) Web-based behavioral modeling for continuous user authentication (CUA), 1st edn, vol 105. Elsevier Inc.
7. Niinuma K, Jain AK (2010) Continuous user authentication using temporal information. In: Biometric technology for human identification VII, vol 7667, p 76670L
8. Niinuma K, Park U, Jain AK (2010) Soft biometric traits for continuous user authentication. IEEE Trans Inf forensics Secur 5(4):771–780
9. Monrose F, Rubin AD (2000) Keystroke dynamics as a biometric for authentication. Futur Gener Comput Syst 16(4):351–359
10. Shen C, Cai Z, Guan X (2012) Continuous authentication for mouse dynamics: a pattern-growth approach. In: Dependable systems and networks (DSN), 2012 42nd annual IEEE/IFIP international conference on, 2012, pp 1–12
11. Altinok A, Turk M (2003) Temporal integration for continuous multimodal biometrics. Multimodal User Authentication 2003(1):11–12
12. Kang H-B, Ju M-H (2006) Multi-modal feature integration for secure authentication. In: International conference on intelligent computing, pp 1191–1200
13. Xie Y et al (2012) Innocent by association : early recognition of legitimate users. In: ACM Conf. Comput. Commun. Secur. (CCS '12), pp 353–364
14. Zhu J, Wu P, Wang X, Zhang J (2013) Sensec: mobile security through passive sensing. In: 2013 international conference on computing, networking and communications (ICNC), pp 1128–1133
15. Saevanee H, Clarke N, Furnell S, Biscione V (2015) Continuous user authentication using multi-modal biometrics. Comput Secur 53:234–246

16. Belk M, Fidas C, Germanakos P, Samaras G (2017) The interplay between humans, technology and user authentication: a cognitive processing perspective. Comput Human Behav 76:184–200
17. Oracle Corporation (2018) Machine learning-based adaptive intelligence : the future of cybersecurity Executive summary, no. January, 2018.
18. Shakir M, Abubakar A, Yousoff O, Waseem M, Al-Emran M (2016) Model of security level classification for data in hybrid cloud computing. J Theor Appl Inf Technol 94(1)
19. Margaret Rouse, Security intelligence (SI). TechTarget network [Online]. https://whatis.techtarget.com/definition/security-intelligence-SI
20. Al-hashimi M, Al-nidawi W-J, Othman M, Shakir M, Sulaiman H (2019) Evaluate information security governance frameworks in cloud computing environment using main and sub criteria, vol 16
21. Shi E, Niu Y, Jakobsson M, Chow R (2010) Implicit authentication through learning user behavior. In: International conference on information security, pp 99–113
22. Hajivali M, Alrashdan MT, Fatemi Moghaddam F, Alothmani AZM (2013) Applying an agent-based user authentication and access control model for cloud servers.In: Int. Conf. ICT Converg., no. October 2013, pp 807–812
23. Shakir M, Abubakar AB, Yousoff Y, Al-Emran M, Hammood M (2016) Application of confidence range algorithm in recognizing user behavior through EPSB in cloud computing. J Theor Appl Inf Technol 94(2)
24. Shakir M (2020) User authentication in public cloud computing through adoption of electronic personal synthesis behavior. Uniten
25. Zhang F, Han D-Y (2012) Applying agents to the data security in cloud computing. In: Computer science and information processing (CSIP), 2012 international conference on, pp 1126–1128
26. Jang E-Y, Kim H-J, Park C-S, Kim J-Y, Lee J (2011) The study on a threat countermeasure of mobile cloud services. J Korea Inst Inf Secur Cryptol 21(1):177–186
27. Suo H, Liu Z, Wan J, Zhou K (2013) Security and privacy in mobile cloud computing. In: Wireless communications and mobile computing conference (IWCMC), 2013 9th international, pp 655–659
28. Bernard L (2011) A risk assessment framework for evaluating software-as-a-service (SaaS) cloud services before adoption. ProQuest Diss. Theses, pp 129-n/a
29. Shen C, Yu T, Xu H, Yang G, Guan X (2016) User practice in password security: an empirical study of real-life passwords in the wild. Comput Secur 61:130–141
30. Dasgupta D, Roy A, Nag A (2016) Toward the design of adaptive selection strategies for multi-factor authentication. Comput Secur 63:85–116
31. Farooq H (2017) A review on cloud computing security using authentication techniques. Int J Adv Res Comput Sci 8(2):19–22
32. Kumar PR, Raj PH, Jelciana P (2018) Exploring data security issues and solutions in cloud computing. Procedia Comput Sci 125(2009):691–697
33. Metheny M (2013) Chapter 9—the FedRAMP cloud computing security requirements
34. Mollah MB, Azad MAK, Vasilakos A (2017) Security and privacy challenges in mobile cloud computing: Survey and way ahead. J Netw Comput Appl 84:38–54
35. Sumitra B, Pethuru CR, Misbahuddin M (2014) A survey of cloud authentication attacks and solution approaches. Int J Innov Res Comput Commun Eng 2(10):6245–6253
36. Tripathi A, Mishra A (2011) Cloud computing security considerations. In: Signal processing, communications and computing (ICSPCC), 2011 IEEE international conference on, pp 1–5
37. Subramanian N, Jeyaraj A (2018) Recent security challenges in cloud computing. Comput Electr Eng 71:28–42
38. Stolen password lets hackers into Deloitte's systems. Industry News [Online]. https://www.securenvoy.com/en-gb/blog/stolen-password-lets-hackers-deloittes-systems

39. Leigh L (2016) Cloud security provider MailGuard partners with Southern Cross Protection to enhance cyber safety. Startup daily., Insights and Stories from the Australian and New Zealand tech ecosystem [Online]. https://www.startupdaily.net/2016/06/mailguard-partnership-sxp/
40. Chen G (2014) Convenience over safety : how authentication cookies compromise user account security on the Web
41. An Y (2012) Security weaknesses of a biometric-based remote user authentication scheme using smart cards, vol 4(3), pp 21–28
42. Aboudagga N, Refaei MT, Eltoweissy M, DaSilva LA, Quisquater J-J (2005) Authentication protocols for ad hoc networks: taxonomy and research issues. In: Proceedings of the 1st ACM international workshop on Quality of service & security in wireless and mobile networks, pp 96–104
43. Niinuma K, Jain A (2010) Continuous user authentication using temporal information. SPIE Defense, Secur. ..., vol 7667, p 11
44. DARPA (2013) Active authentication. Authenticatio. [Online]. http://www.darpa.mil/program/active
45. Hao Z, Zhong S, Yu N (2011) A time-bound ticket-based mutual authentication scheme for cloud computing. Int J Comput Commun Control 6(2):227–235
46. Omri F, Hamila R, Foufou S, Jarraya M (2012) Cloud-ready biometric system for mobile security access. In: International conference on networked digital technologies, pp 192–200
47. Le Z, Zhang X, Gao Z (2013) NemoAuth: a mnemonic multimodal approach to mobile user authentication. In: TENCON 2013–2013 IEEE region 10 conference (31194), pp 1–6
48. Banyal RK, Jain P, Jain VK (2013) Multi-factor authentication framework for cloud computing. In: Proc. Int. Conf. Comput. Intell. Model. Simul., pp 105–110
49. Yang YH, and P. Y. Lin, "An ID-based user authentication scheme for cloud computing," in *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2014 Tenth International Conference on*, 2014, pp. 98–101
50. Cindhamani J, Punya N, Ealaruvi R, Dhinesh LD (2014) An enhanced data security and trust management enabled framework for cloud computing systems
51. Al Rassan I, AlShaher H (2014) Securing mobile cloud computing using biometric authentication (SMCBA). In: Computational science and computational intelligence (CSCI), 2014 international conference on, vol 1, pp 157–161
52. Chang V, Kuo Y-H, Ramachandran M (2016) Cloud computing adoption framework: a security framework for business clouds. Futur Gener Comput Syst 57:24–41
53. Dey S, Sampalli S, Ye Q (2013) Message digest as authentication entity for mobile cloud computing. In: Performance computing and communications conference (IPCCC), 2013 IEEE 32nd international, pp 1–6
54. Grzonkowski S (2010) Sedici: an authentication service taking advantage of zero-knowledge proofs. In: International conference on financial cryptography and data security, p 426
55. Jayawardana V (2017) Adaptive authentication and machine learning. Towards Data Science [Online]. https://towardsdatascience.com/adaptive-authentication-and-machine-learning-1b460ae53d84
56. Lim SY, Mat Kiah ML, Ang TF (2017) Security issues and future challenges of cloud service authentication. Acta Polytech Hungarica 14(2):69–89
57. Chow R et al (2010) Authentication in the clouds: a framework and its application to mobile users. In: Proceedings of the 2010 ACM workshop on Cloud computing security workshop, pp 1–6
58. Song Z, Molina J, Lee S, Lee H, Kotani S, Masuoka R (2009) Trustcube: an infrastructure that builds trust in client. In: Future of trust in computing. Springer, pp 68–79
59. Kang H-B, Ju M-H (2006) Multi-modal feature integration for secure authentication. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol 4113 LNCS, pp 1191–1200
60. Barkadehi MH, Nilashi M, Ibrahim O, Zakeri Fardi A, Samad S (2018) Authentication systems: a literature review and classification. Telematics Inform 35(5):1491–1511

61. Oeltjen J, Authentication and machine learning: taking behavior recognition to a new level. RSA Identity Reimagined. [Online]. https://www.csoonline.com/article/3209917/identity-management/article.html
62. Kimlaychuk V (2013) Authentication using shared knowledge: learning agents. Adv Intell Syst Comput 194 AISC 2:523–531
63. Information F, Standards P (2004) Standards for security categorization of federal information and information systems, no. February, 2004
64. Bolle RM, Connell JH, Ratha NK (2002) Biometric perils and patches. Pattern Recognit 35 (12):2727–2738
65. Golfarelli M, Maio D, Malton D (1997) On the error-reject trade-off in biometric verification systems. IEEE Trans Pattern Anal Mach Intell 19(7):786–796
66. Ko RKL et al (2011) TrustCloud: a framework for accountability and trust in cloud computing. In: Services (SERVICES), 2011 IEEE World Congress on, 2011, pp 584–588
67. Farcasin M, Guli A, Chan-Tin E (2017) Fluid passwords-mitigating the effects of password leaks at the user level. arXiv Prepr. arXiv1708.09333
68. Jefferson D, Rubin AD, Simons B, Wagner D (2004) Analyzing internet voting security. Commun ACM 47(10):59–64
69. Subashini S, Kavitha V (2011) A survey on security issues in service delivery models of cloud computing. J Netw Comput Appl 34(1):1–11
70. Nikam R, Potey M (2017) Cloud storage security using Multi-Factor Authentication. In: 2016 Int. Conf. Recent Adv. Innov. Eng. ICRAIE 2016
71. Chen T-H, Yeh H, Shih W-K (2011) An advanced ecc dynamic id-based remote mutual authentication scheme for cloud computing. In: Multimedia and ubiquitous engineering (MUE), 2011 5th FTRA international conference on, pp 155–159
72. Odelu V, Das AK, Kumari S, Huang X, Wazid M (2017) Provably secure authenticated key agreement scheme for distributed mobile cloud computing services. Futur Gener Comput Syst 68:74–88
73. Shepherd J (2016) Duo + OneLogin: Adaptive Authentication. Product and technology, security & compliance [Online]. https://www.onelogin.com/blog/what-is-adaptive-authentication
74. Paper W, Paper W, Adaptive authentication superior user experience and growth through intelligent
75. Ferdowsi A, Saad W (2018) Deep learning for signal authentication and security in massive internet of things systems, pp 1–30
76. Altinok A, Turk M (2003) Temporal integration for continuous multimodal biometrics. In: Proceedings of the workshop on multimodal user authentication, no. 1
77. Fisher N (2018) 5 Identity attacks that exploit your broken authentication. Product Marketing Manager, Security, 2018. [Online]. https://www.okta.com/security-blog/2018/03/5-identity-attacks-that-exploit-your-broken-authentication/%0D. Accessed: 06-Oct-2018
78. Eldefrawy MH, Al-Muhtadi JF (2016) Cryptanalysis and enhancement of a password-based authentication scheme. In: Proc. - IEEE 7th Int. Conf. Cloud Comput. Technol. Sci. CloudCom 2015, pp 548–551
79. Kaushik A, Awashti HO, Goel K, Goel S (2012) Secure authentication with encryption technique for mobile on cloud computing 1(5):28–33
80. Bennett A (2018) 8 Public cloud security threats to enterprises in 2018. DELL EMC, IT security, networking and cloud technology, 2018. [Online]. https://www.comparethecloud.net/articles/8-public-cloud-security-threats-to-enterprises-in-2017/. Accessed: 06-Oct-2018
81. Fuller C (2018) Complying with NIST guidelines for stolen passwords. Security Bloggers Network. [Online]. https://securityboulevard.com/2018/03/complying-with-nist-guidelines-for-stolen-passwords/. Accessed: 06-Oct-2018
82. Ruoti S, Andersen J, Seamons K (2016) Strengthening Password-based Authentication. In: Twelfth symp. usable priv. secur. (SOUPS 2016)
83. Killoran J (2018) 4 password authentication vulnerabilities & how to avoid them. Swoop founders and Matt Custer

An Intrusion Detection System Based on Hybrid of Particle Swarm Optimization (PSO) and Magnetic Optimization Algorithm (MOA)

Siti Norwahidayah Wahab, Noor Suhana Sulaiman, Noraniah Abdul Aziz, Nur Liyana Zakaria, and Ainal Amirah Abd Aziz

Abstract IDS (Intrusion Detection System) is a security component that protects computer and network systems. Variety of methods have been developed to improve the IDS accuracy. One of the most recent approaches is the use of real-time monitoring and irregular activity detection. When an intrusion is detected, a message will be sent to the network administrator. One of the disadvantages of IDS is the possibility of a bad packet by-passing through network traffic. As a result, an improvement on the Artificial Neural Network (ANN) is explored in this study to enhance attack detection in IDS. Standard and attack events are described using the NSL-KDD dataset. In this study, the Magnetic Optimization Algorithm (MOA) is combined with Particle Swarm Optimization (PSO) named PSOMOA, thus to increase the classification rate and achieve high detection accuracy in IDS. MOA is a heuristic optimization algorithm which deals with attraction between particles scattered in the search space and inspired by magnetic field theory in physics. NSL-KDD dataset represented as attacks and normal activities used in this study. Smurf and Neptune attacks are selected for testifying detection and classification ability of attack category of proposed PSOMOA. During the experimentation process, four of the most important features of the dataset were selected. The PSOMOA findings are compared to those of the other form, which employs PSO and MOA. According to the obtained results, the proposed PSOMOA could increase IDS accuracy by up to 99.5%.

Keywords Artificial neural network · Magnetic optimization algorithm · Particle swarm optimization · Intrusion detection system · NSL-KDD

S. N. Wahab (✉) · N. S. Sulaiman · N. A. Aziz · N. L. Zakaria · A. A. A. Aziz
Faculty of Computer, Media and Technology Management, University College TATI, Terengganu, Malaysia
e-mail: norwahidayah@uctati.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_46

1 Introduction

Intrusion Detection System (IDS) protects the network from malicious activity. An intrusion detection system may also be thought of as "powerful security elements capable of detecting, preventing, and responding to computer attacks" [1]. Recent technology has been used to track and identify abnormal behavior in real time. When an intrusion is observed, the IDS notifies and sends alerts to the network administrator. Until now, IDS was used throughout the network to detect an intrusion by monitoring network traffic. Intrusion detection has previously relied on evaluating network flows, logs, and device events [2]. Deploying an IDS in a large-scale network, on the other hand, is a challenging task since it exposes the network to a range of attacks. The problems and drawbacks of IDS do not justify a business or company's refusal to use it. This research has emphasized the importance of studying and improving the best way to solve those issues.

Many analysts nowadays use a modern methodology to address IDS's nagging issues. Artificial Intelligence (AI) is the study of the possibility of creating intelligent systems that can reason and think like humans. IDS employs a variety of AI techniques. The Artificial Neural Network (ANN) can solve classification and regression problems. Its ability to learn from observing the dataset, however, has some limitations. On the other learning methods, Feedforward Neural Network (FNN), aim to find the best combination of connection weight and biases to achieve the lowest possible error. FNN frequently converge to points that are the best local solution but not globally suitable for large network. Simulated Annealing (SA), Genetic Algorithm, and other heuristic optimization methods for FNN have been suggested by several researchers. Genetic algorithm (GA), Particle Swarm Optimization (PSO), Magnetic Optimization Algorithm (MOA), and Differential Evaluation (DE) are number of examples of optimization algorithms [3]. According to Mirjalili et al. [3], it has some disadvantages over other optimization approaches, such as slow convergence speeds.

Denial of Service (DoS) attacks degrade server output by flooding ICMP traffic, and they have been happening since February 2000 [4]. Aside from that, DoS will cause a victim's server memory resources to become full by repeated sending TCP packet for requesting to start a TCP session [4]. Smurf attack and Neptune attack, both are kind of DoS attacks. As such, detecting DoS attacks is an essential and required tool and technique to employ, particularly on Smurf and Neptune attack. As a result, a combination of Partcle Swarm 4 Optimization (PSO) and Magnetic Optimization Algorithm (MOA) called hybrid PSOMOA was proposed to solve the ANN drawbacks.

2 Literature Review

In network analysis, network datasets are often used. Large amounts of time are spent processing and transmitting data, as well as a higher rate of false alarms in attack detection, all of which result in harm and unsafe conditions for online users, such prescribed in IDS dataset [2].

Researchers have suggested several optimization strategies to achieve high precision in intrusion detection [5]. Previous work and methods will be highlighted in this topic, such as the hierarchical clustering algorithm [6], removal feature selection and support vector machine [7], gravitational search algorithm [5], and pattern recognition (Ernesto et al. 2015).

Horng et al. [6] proposed a combination of hierarchical clustering and support vector machine (SVM), in which the author used the (Balanced Iterative Reducing and Clustering using Hierarchies) Birch algorithm to reduce the KDD CUP 1999 dataset. Before dataset training, the SVM, an algorithm created a reduced and high-quality dataset from the original KDD CUP 1999 dataset. The Birch algorithm is an unsupervised data mining algorithm for performing hierarchical gathering over large datasets [6]. The experimentation resulted in a high accuracy of 95.72% and a low false positive rate of 0.7% using the proposed techniques.

The other method is Gradual Feature Removal (GFR) [7]. The paper uses 19 important features to represent a network visit [7]. To achieve better results, the author also used a hybrid feature selection approach that combines wrapper and filter methods. The first stage is data pre-processing, which involves creating a raw database and removing duplicate data. The dataset was then trained using the Artificial Neural Network (ANN) process, which classified the data using 41 KDD CUP'99 features to find the best training sample with high accuracy. The final stage is feature reduction, which involves training and testing the dataset with 41 features and four separate feature reduction techniques before the author obtains 19 essential features using the GFR process.

Dastanpour et al. [5] proposed the Gravitational Search Algorithm (GSA) to assist ANN in IDS. The GSA is a popular machine learning algorithm based on gravity law and mass interaction. The four conditions of mass, referred to as agents, are location, inertial mass, active and passive gravitational mass. The author divides the KDD CUP'99 dataset into two phases which for testing and training. After that, the methods for reorganization established a structured format for datasets. When the ANN training is complete, the ANN classifies the testing dataset and outputs the accuracy of the device detection as a result. Following the acknowledgment of the ANN end, the GSA optimizes the ANN. The outcome of ANN (without the GSA) will be compared to the result of ANN with the GSA (with GSA). For the dataset, author decide to set 1 represent for the attack and 0 as normal for separate the attack and normal one. The experimental result show that the ANN with GSA contributes high accuracy (98.7%) rate of detection. Proposed method capable to optimize and improve the ANN performance. Besides, the method also achieves 1.0 accuracy with only 39 critical features of KDD CUP'99 [5].

Tian and Liu [8] proposed a procedure that uses an ANN and a PSO algorithm (PSOA). PSOA was proposed by the authors because it can address the problem of searching performance. The paper used a rough collection of ANN data to select a subset of input attributes and used PSOA to improve ANN efficiency. The authors used an equation to determine each fitness particles, update velocity and position, and update local best and global best particle to integrate the PSOA [8].

Eight rough set features were chosen, with six nodes in the ANN output layer. One of the five output nodes was used for standard, while the other four were used for attacks. The training set is 80% and testing set is 20% where the total of the data is 460. The proposed method has a higher level of stability and can detect more attacks with a lower mean square error. The PSOA, on the other hand, has some flaws, including a slow convergence rate [3].

3 NSL-KDD Dataset

NSL-KDD dataset is offline network data based on KDD 99 dataset where it contains 41 attributes and one class attribute. NSL-KDD is a data set suggested to solve some of the inherent problems of the KDD'99 data set where it eliminates noise data in KDD'99 data set. The NSL-KDD dataset is smaller than the KDD99 dataset which has duplicate records. The advantages of NSL-KDD data set are:

(i) The train collection does not contain any duplicate records, so the classifiers would not favor more frequent records.
(ii) Since the proposed test sets contain no duplicate records, methods with higher detection rates on frequent records have no impact on the learners' performance.
(iii) The number of records selected from each difficulty level group is inversely proportional to the percentage of records in the original KDD data collection.
(iv) As a result, different machine learning methods' classification rates vary more widely, making an accurate evaluation of different learning techniques more reliable.
(v) Since the train and test sets contain a reasonable number of records, it is more practical to test the entire set rather than a small subset at random.

NSL-KDD dataset divided into two sets which are training sets and testing sets. Each record of the dataset is continuous, discrete and symbolic forms. Protocol_type feature that used in this study represent 2 as tcp, 3 as udp, 4 as icmp and 5 as other symbols. Besides, for the result used 0 for Normal and 1 as Attack. The attacks can be categorized as follows:

(i) Denial of Service Attack (DoS): DoS attack is an aggressive attack that an attacker sends many ping packet to make the network flooded with packet and become congested. Such that, the attacker can denies the legitimate user access to the servers (Mahbod et al. 2009).

(ii) User to Root Attack (U2R): The attacker has a local access to the victim machine and to gain the root access to the system [9].
(iii) Remote to Local Attack (R2L): The attacker did not have a local access to the victim machine but tries to gain the local access as a user to the machine [9].
(iv) Probe Attack: Intruders try to grab information about the target machine and aim to bypass the security controls (Mahbod et al. 2009). Generally, most researchers used NSL-KDD which include the total of dataset is 494,020. The distribution dataset for Normal is 97280, Probe is 4107, DoS is 391458, U2R is 52, R2L is 1124 [10].

4 Methodology

The methodology for this research is mainly divided into five sections; in which the first part will be on collecting and processing dataset to be used for training ANN in IDS. The second part will focus on applying the hybrid PSOMOA on dataset. The dataset will then also be used to classify the normal and abnormal activities. The output value of PSOMOA will support ANN to maximize the classification rate. The final parts involved the analysis and comparison of the optimization by using ANN supported by PSOMOA.

4.1 Pre-Processing

There are 41 attributes and one class attribute in the NSL-KDD data set. Some of the 41 attributes play no role in attack detection, whereas others play a minor role. The knowledge gain attribute evaluation, gain ratio attribute evaluation, and correlation attribute evaluation algorithms are used by Bajaj et al. [11]. According to the researcher [11], attributes 9, 20, and 21 play no part, although attributes 15, 17, 19, 32, 40 have minimum role in detection of attack. An analysis of the NSL-KDD dataset reveals that features 7, 8, 11, and 14 have virtually all zero values. By eliminating these least available features from the dataset's training and testing sets, the dataset would be reduced to 29 features, resulting in a smaller dataset. Now that the dataset has been limited, it can be used for training and research. A total of 41 features were used in the training and testing, and feature reduction is not performed on dataset.

It is impossible to use all rows of dataset because it takes a lot of run time. Hence, in this study, only 4 critical dataset features selected and 200 rows input

values will be used to train ANN. Besides, one column will be added for identifying each record as a result which are 0 for normal and 1 for attack in the standard format step. The reason of adding this additional row is to continue and understand the error reorganization in ANN. In order to make the dataset compatible, the given attributes are converted to a double data form. with the ANN toolbox of MATLAB toolkit because it is only support the data in integer form [12]. The "protocol_type" feature converted with values like tcp is 2 and udp is 3. Not all protocol types selected because only 200 rows of dataset used to test the algorithm [12].

In this research, the MOA used to optimize the results of ANN. MOA makes the random agent calculate the mass value of each agent to optimize the ANN. The MOA parameters that used in this study is number of agents, maximum number of iteration, number of training samples, inertia weight, minimum weight, maximum weight and objective function.

4.2 Propose IDS Security System Based on Hybrid PSOMOA

The method proposed in this paper is for training ANN to optimize IDS with supported by the combination of PSO and MOA. First, this study tries to divide the dataset into two data section for training and testing. Then the methods try to develop the standard dataset format for the reorganization of ANN. After the training of the ANN is completed, the ANN will classify the NSL-KDD testing dataset and extract the accuracy output of the system detection, which will then be plotted and monitored it by the system. When the recognition phase of the ANN is completed, then the MOA data input will be the result of ANN. At this stage, the ANN reorganization will attempt to be optimized by MOA. Finally, after the results of the ANN are optimized by the MOA, they will by plotted to be compared with the ANN result (without MOA), such that the effects of MOA will be more understood in the ANN reorganization in the intrusion detection system using the dataset of NSL-KDD. The main idea and overall methods have been illustrated in Fig. 1. In order to support ANN in IDS optimization, the PSO will be implemented.

The training method involve once packet data is extracted. The useless data are removed. The knowledge is any keep as patterns to form a knowledge file. The information file contains immense patterns. The patterns square measure any processed to extract representative patterns which will be used for coaching the ANN with specific ANN rule. The result is a final weight matrix, which is saved in a large format. Incoming packets of data knowledge are stripped and any tangential information is filtered as part of the testing process. The data is then processed with the ultimate weights to produce a price in the ANN's output layer. The assisted performance is based on intrusion detection classification and testing is completed.

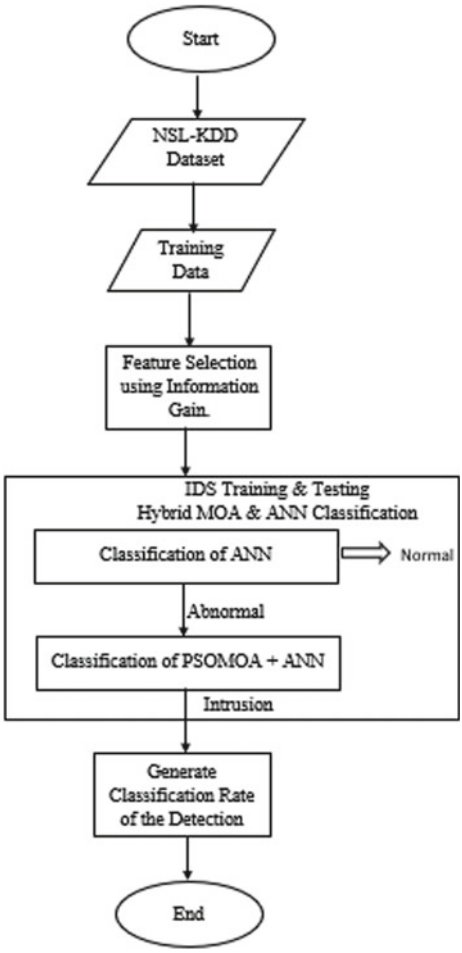


Fig. 1 Process flow

5 Result and Discussion

In this research, only 4 critical features selected to classify the attack and normal packet in IDS. The experimentation will compare the classification rate result with another best features selection. Another relevant features selected include 3, 5, 29 and 39 which 3 for Service (network service on the destination such as http, udp and etc.), 5 for Source Bytes (number of data bytes from source to destination), 29 for Same Srv Rate (percentage of connection to the same service) and 39 for Dst Host Srv Serror Rate (Service error rate for the destination host) [13].

Besides, the other relevant dataset features is number 3, 23, 30 and 36 [14]. Feature 3 for Service (network service on the destination such as http, udp, etc.), 23 is Count (number of connections to the same host as the current connection in the past two seconds), 30 for Diff Srv Rate (percentage of connections to the different

services) and 36 for Dst Host Same Src Port Rate (same source port rate for destination host). The next relevant features from [13] are 36, 37, 38 and 39 where 36 for Dst Host Same Src Port Rate (same source port rate for destination host), 37 is Dst Host Srv Diff Host Rate (Different host rate for destination host), 38 is Dst Host Serror Rate (error rate for destination host) and 39 is Dst Host Srv Serror Rate (service error rate for destination host).

The last relevant dataset features selected are 3, 6, 29 and 30 [13]. Feature 3 is Service (network service on the destination such as http, udp, etc.), 6 is Dst Bytes (number of data bytes from destination to source), 29 is Same Srv Rate (percentage of connections to the same services) and 30 is Diff Srv Rate (percentage of connections to different services). In this study, all combination the dataset features will be simulated and produce up to total result of 37 different classification rate. Table 1 shows the relevant features of detecting the normal behavior, Smurf attack and Neptune attack.

This study used most of the relevant features, to detect those attacks in IDS. Table 2 list the combination of relevant features to investigate to get the best features selection. The dataset will load in the MATLAB using the code of MOA, PSO and Hybrid of MOA and PSO to produce the high classification rate of the attack detection.

In Fig. 2 shows, FNN using hybrid PSO and MOA (so called PSOMOA) is the better techniques for optimization on ANN that can achieve high accuracy compared with others. The parameter setting that suitable in this project is using masses 30, 500 for maximum number of iterations, 15 for hidden nodes, 200 training

Table 1 Relevant features for Normal, Smurf attack and Neptune attack

Class label	Relevant features
Normal	3, 6, 12, 23, 25, 26, 29, 30, 33, 34, 35, 36, 37, 38, 39
Smurf	2, 3, 5, 6, 12, 25, 29, 30, 32, 36, 37, 39
Neptune	3, 4, 5, 23, 26, 29, 30, 31, 32, 34, 36, 37, 38, 39

Table 2 Combination of relevant features of Normal, Smurf Attack and Neptune Attack

Class label	Features	Features name
Normal	2, 3, 4, 5, 6, 12, 23	Protocol_type, Service, Flag, Src_bytes, Dst_bytes, Logged_in, Count, Serror_rate, Srv_serror_rate
Smurf	25, 26, 29, 30, 31 32, 33, 34, 35, 36	Same_srv_rate, Diff_srv_rate Srv_diff_host_rateDst_host_count
Neptune	37, 38, 39	Dst_host_srv_count, Dst_host_same_srv_rate, Dst_host_diff_srv_rate, Dst_host_same_src_port_rate, Dst_host_srv_diff_host_rate, Dst_host_serror_rate, Dst_host_srv_serror_rate

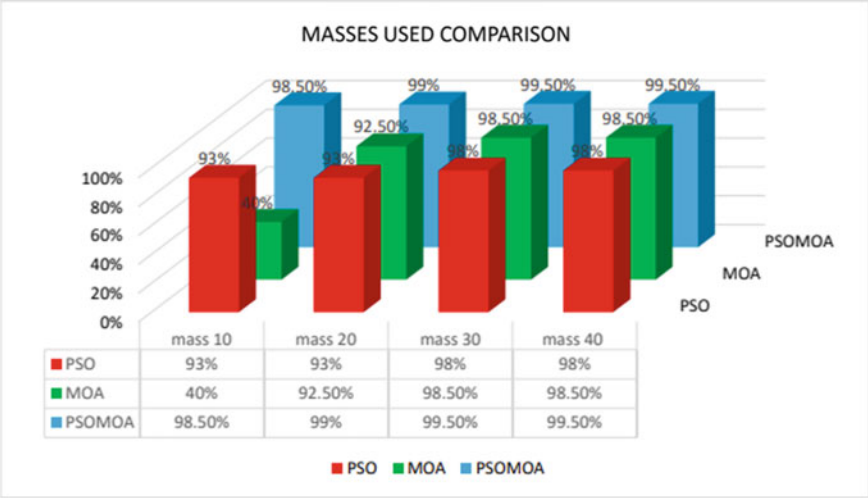


Fig. 2 Comparison of Masses 10, 20, 30 and 40 of MOA, PSO and PSOMOA

samples and 2 for inertia weight. Using this parameter setting with the optimization algorithm proposed, it can achieve high detection rate and high classification rate with low iteration numbers and masses. According to the result given, PSOMOA achieves the reliability because it reduces the probability of trapping in local minima.

PSOMOA's outcome demonstrates that it is capable of both searching and attracting people. In other words, the strengths of PSO and MOA have been successfully combined to produce excellent results in FNN preparation.

6 Conclusion

A new training optimization algorithm which are hybrid of ANN, PSO and MOA is introduced and investigated. NSL-KDD dataset used to represent a packet and simulated in the MATLAB toolkit. The project steps proposed as follows: collecting and processing dataset which select only 4 critical features for detection of two network attack categories to be used for training ANN in IDS; applying the MOA on dataset; classify the dataset which are normal and abnormal activities; output value of MOA will support ANN to maximize the classification rate and the final parts involved the analysis and comparison of the optimization by using ANN supported by MOA.

This project uses the Magnetic Optimization Algorithm (MOA) in combination with Particle Swarm Optimization (PSO) to determine the accuracy of Intrusion Detection System (IDS). The simulation method applied in MATLAB toolkit

yielded a comparison of classification rates using MOA, PSO, and a combination of PSO and MOA. Those optimization approaches lead to different outcomes.

As a conclusion, the hybrid of PSOMOA produce highest accuracy and classification rate for high numbers of training samples. The combination optimization method proven that the MOA can support ANN and PSOMOA is a reliable technique which perform high accuracy and efficiency due to its function like a magnetic that can attract more particles. The classification rate of hybrid of PSOMOA achieves 99.5%, PSO achieves 98% and MOA achieves 98.5%.

Acknowledgements This research is fully supported by Fundamental Research Grant Scheme (FRGS), which help in completing the research in viable and efficient.

References

1. Kumar G, Kumar K (2012) The use of artificial-intelligence-based ensembles for intrusion detection: a review. Hindawi Publishing Corporation
2. Sulaiman NS, Aziz NS, Samsudin N, Mohamed WAAW (2020) Big data analytic of intrusion detection system. Int J Adv Trends Comput Sci Eng
3. Mirjalili SA, Hashim SZM, Sardroudi HM (2012) Training feedforward neural networks using hybrid particle swarm optimization and gravitational search algorithm. Elsevier
4. Solankar P, Pingale S, Parihar RS (2015) Denial of service attack and classification techniques for attack detection. (IJCSIT) Int J Comput Sci Inf Technol 6(2)
5. Dastanpour A, Ibrahim S, Mashinchi R, Selamat A (2014) Using gravitational search algorithm to support artificial neural network in intrusion detection system. Smar Comput Rev 4(6)
6. Horng S-J, Su M-Y, Chen Y-H, Kao T-W, Chen R-J, Lai J-L, Perkasa CD (2010) A novel intrusion detection system based on hierarchical clustering and support vector machines. Elsevier
7. Li Y, Xia J, Zhang S, Yan J, Ai X, Dai K (2011) An efficient intrusion system based on support vector machines and gradually feature removal method. Elsevier
8. Tian WJ, Liu JC (2010) A new network intrusion detection identification model research. Int Asia Conf Inf Control Autom Robot
9. Paliwal S, Gupta R (2012) Denial-of-service, probing & remote to user (R2L) attack detection using genetic algorithm. Int J Comput Appl 60(19)
10. Aggarwal M, Amrita (2013) Performance analysis of different feature selection methods in intrusion detection. Int J Sci Technol Res 2(6)
11. Moradi M, Zulkernine M (2004) A neural network based system for intrusion detection and classification of attacks. In: Natural Sciences and Engineering Research Council of Canada (NSERC)
12. Mukhopadhyay I, Chakraborty M (2014) Hardware realization of artificial neural network based intrusion detection & prevention system. J Inf Secur
13. Olusola AA, Oladele AS, Abosede DO (2010) Analysis of KDD "99 intrusion detection dataset for selection of relevance features. In: Proceedings of the world congress on engineering and computer science, vol I
14. Chandolikar NS, Nandavadekar VD (2012) Selection of relevant feature for intrusion attack classification by analyzing KDD CUP 99. MIT Int J Comput Sci Inf Technol 2(2)
15. Sulaiman NS, Aziz NS, Samsudin N, Alyani WA, Yacob A, Ngah L (2020) Overview of network dataset and data mining technique. Int J Adv Trends Comput Sci Eng

Medical Informatics

An Enhancement of Succinate Production Using a Hybrid of Bacterial Foraging Optimization Algorithm

Shen Yee Siow, Mohd Saberi Mohamad, Yee Wen Choon, Muhammad Akmal Remli, and Hairudin Abdul Majid

Abstract Genetic modifications, such as gene knockout technique, have become mainstream in metabolic engineering to produce desired amount of targeted metabolites through reconstruction of the metabolic networks. The production, however, does not often achieve desirable outcome. To this end, in-silico methods have been applied to predict potential metabolic network response and optimise production. Previous methods working on relational modelling framework, such as OptKnock and OptGene, however, failed at handling its multivariable and multi-modal functions optimization algorithms. This paper proposes hybridising bacterial foraging optimizationg algorithm (BFO) and dynamic flux balance analysis (DFBA) to overcome problems in OptKnock and OptGene with a nature-inspired algorithm and also to couple kinematic variables in the model to predict production of succinate in *E.coli* model. In-silico results showed that by knocking out genes identifed by BFODFBA, production rate of succinate is better as when compared to OptKnock and OptGene.

Keywords Bacterial foraging algorithm · Dynamic flux balance analysis · Gene knockout · Metabolic engineering · *Escherichia coli*

S. Y. Siow · H. A. Majid
Artificial Intelligence and Bioinformatics Research Group, School of Computing, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Skudai, Malaysia
e-mail: hairudin@utm.my

M. S. Mohamad (✉)
Health Data Science Lab, Department of Genetics and Genomics, College of Medical and Health Sciences, United Arab Emirates University, P.O. Box 17666, Al Ain, Abu Dhabi, United Arab Emirates
e-mail: saberi@uaeu.ac.ae

Y. W. Choon · M. A. Remli
Institute for Artificial Intelligence and Big Data, Universiti Malaysia Kelantan, Kota Bharu 16100, Kelantan, Malaysia
e-mail: akmal@umk.edu.my

Y. W. Choon · M. A. Remli
Department of Data Science, Universiti Malaysia Kelantan, Kota Bharu 16100, Kelantan, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_47

1 Introduction

Gene knockout strategy works at genetically modifying genes of an organisms to render them inoperative or being removed. This is a widely used technique especially in the production of biochemical substances.

Succinate are very important chemicals in various fields. Succinate and its byproducts are important chemicals. They are widely used in various fields such as in pharmaceuticals, industrial and even environmentally friendly biodegradable polyester. In addition, succinate and its byproducts have also become highly demanded in food processing industries. However, traditionally produced succinate through petrochemical processes failed at meeting demands. Yields from these traditional methods are significantly less than the theoretical maximum due to the absence of metabolic and genetic engineering interventions.

In-vitro experimentation to increase production of succinate through petrochemical process lacks cost-effectiveness, is time-consuming and can cause pollution. The use of microorganism has become mainstream in producing succinate. *E coli* breaks down glucose and produce succinate. It has been chosen as a production host, over other microorganisms, because of it many flexible robust characteristics such as growth in anarobic and aerobic conditions and simple nutrients requirements. As such, in-silico methods of intervening with the production of succinate in *E. coli* was able to be introduced [1].

Previous gene knockout algorithms such as OptKnock [2] and its extension OptReg [3] and RobustKnock [4], were used to identify reactions in *E. coli* to be modified in order for the organism to overproduce succinate. Also, there were some hybrid algorithms that had been implemented to perform gene knockout to enhance the production of certain metabolites. But the production still does not meet a desirable result due to due to the difficulties specifically in the performance of multivariate and multimodal functions and unrealistic flux distributions from lack of regulatory kinetic data in *E. coli* model.

Hence, this paper aims at extending the applicability of OptKnock by hybridizing nature-inspired algorithms such as BFO and DFBA, neither of which deal with identification of ideal knockout gene set independently. BFO will identify ideal genes to be knocked out and subsequently DFBA will be used to predict growth rate of *E. coli* and production of succinate after the knockout.

2 Materials and Methods

2.1 Datasets and Experimental Setup

Dataset used in this research is the *E. coli* core model, which is in the.mat format. This model can be downloaded from the website of System Biology Research Group (http://systemsbiology.ucsd.edu/Downloads/EcoliCore). It is a subset of the

*i*AF1260 model after modification. The *E. coli* core model consists of 134 genes, 95 reactions, and 72 metabolites. In this paper, the focus is on the production of succinate caused by deletion of certain reactions inside the cell, the fields of data required from the dataset model are *rxns*, *rxnNames*, *mets*, *metNames*, *genes*, *lb* and *ub*, which represent reactions pathway, reaction name, metabolites, metabolite names, genes involved, lower boundary and upper boundary, respectively.

Table 1 shows the parameter values for BFODFBA. The parameters are determined after a number of trial-and-error while some of the parameters are set based on the default value from the original author.

2.2 BFO Algorithm

Evolution and natural genetics-based optimization algorithms have been widely used since its introduction five decades ago, including the more recent and more efficient swarm-based algorithms, BFO [5]. BFO was widely recognized as a global optimization algorithm of current interest for distributed optimization and control and has also attracted the interest of researchers for its effectiveness in tackling real-world optimization challenges in a variety of application domains. BFO is inspired by food hunting behaviour of E coli in human intestine, chemotaxis. Due to its ease of use and flexibility, BFO has been applied effectively applied to solve many engineering problems. BFO was also paired with other optimization approaches. The most recent work is proposed by Chen [6] on an enhanced BFO with chaotic chemotaxis step length, Gaussian mutation and chaotic local search.

As the BFO algorithm is motivated by the hunting actions of *E. coli*, it follows exactly the *E. coli* bacteria. This method is used for locating, handling and ingesting the food. During searching, two actions are performed by the bacterium, which are tumbling or swimming. Tumbling causes the bacterium to change orientation while during swimming (or chemotaxis), the bacterium continues to move in its original direction [7]. Foraging strategy of E. coli bacteria is controlled by four processes:

1. Chemotaxis
2. Reproduction

Table 1 Parameters used in the developed hybrid algorithm, BFODFBA

Parameter	Value
Number of iterations (iteration)	100
Number of populations (Npopulation)	100
Number of bacteria (NBac)	100
Number of chemotactic steps (Nc) *	2
Length of a swim (Ns) *	1
Number of reproduction (Nre) *	2
Number of elimination (Ned) *	1
Number of gene knockouts (20 runs each)	2, 3, 4, and 5

* represents the default value from original author

3. Elimination and dispersal
4. Swarming.

2.3 Dynamic Flux Balance Analysis

Flux Balance Analysis (FBA) traditionally analyses microbial metabolic network, specifically metabolic flux at steady state. However, metabolic engineering involves many reprogramming of the metabolic network, hence, the dynamics of which must also be analysed, limiting greatly the use of FBA in dynamic metabolic flux. The dynamics of these metabolic networks must also be analysed. Limitation faced by FBA such as correctly predicting metabolite concentration in dynamic flux environment can be solved by implementing nonlinear and linear programming of short time intervals, as proposed in Dynamic FBA (DFBA). DFBA can forecast the dynamics of diauxic growth. DFBA also makes the incorporation of kinetics if kinetics is well categorized.

A major improvement of DFBA from FBA is its DOA and SOA approaches. Both can solve dynamic optimization problem by first breaking down flux profiles and its metabolite ratio into discreet time interval and solve the problem of instantaneous optimization and then integration over the whole-time span of interest. The implementation of nonlinear programming (NLP) in DFBA is a major development and will likely to find many applications in computational sciences and biotechnological studies in the future. On the other hand, SOA involves dividing the batch time into several time intervals and solving the problem of instant optimization at the beginning of each time interval, followed by integration over the interval. During the batch phase, the optimization problem was repeatedly solved by using LP to achieve a flux distribution at a given time. DFBA is a significant development of the classical FBA and can be used as a computational analysis tool for fundamental sciences and biotechnology [8]. Equation (1) shows the formulation of a general dynamic optimization problem.

$$
\begin{aligned}
&\underset{\mathbf{z}(t),\mathbf{v}(t),X(t)}{\text{Max}}\ \hat{w}_{\text{end}}\Phi(\mathbf{z},\mathbf{v},X)|_{\text{t=t}_\text{f}} + \hat{w}_{\text{ins}}\sum_{\text{j=0}}^{\text{M}}\int_{\text{t}_0}^{\text{t}_\text{f}} L(\mathbf{z},\mathbf{v},X(t))\delta(t-t_\text{j})\text{d}t \\
&\text{s.t.}\ \frac{\text{d}\mathbf{z}}{\text{d}t} = A\mathbf{v}X \\
&\quad \frac{\text{d}X}{\text{d}t} = \mu X \\
&\quad \mu = \Sigma w_i v_i \\
&\quad t_j = t_0 + j\frac{t_\text{f}-t_0}{M} \quad j = 0\cdots M \\
&\quad c(\mathbf{v},\mathbf{z}) \le 0 \quad |\dot{\mathbf{v}}| \le \dot{\mathbf{v}}_{\max} \quad \forall t \in [t_0, t_\text{f}] \\
&\quad \mathbf{z} \ge 0 \quad X \ge 0 \quad \forall t \in [t_0, t_\text{f}] \\
&\quad \mathbf{z}(t_0) = \mathbf{z}_0 \quad X(t_0) = X_0,
\end{aligned}
\tag{1}
$$

where z_0 and X_0 are the initial conditions for the concentration of the metabolite and the concentration of biomass, $c(\boldsymbol{v}, \boldsymbol{z})$ is the vector function describing nonlinear restrictions that may exist due to the consideration of the kinetic expressions for the fluxes, t_0 and t_f are the initial and final periods, ϕ is the terminal objective function that depends on the concentration of the endpoint, L is the final objective function. δ is Dirac-delta function, t_j is the time instant at which L is considered, $\hat{w}_{ins}$ and $\hat{w}_{end}$ are the weights correlated with instantaneous and terminal objective function, and $\boldsymbol{v}(t)$ is the time profile of metabolic fluxes. If there is no nonlinear restriction, the problem can be reduced to an algorithm involving a bilinear system.

2.4 A Hybrid of BFO and DFBA

A hybrid of BFO and DFBA (BFODFBA) is suggested in this paper in order to resolve some issues from previous researches such as unrealistic knockout strategies are produced and the lack of dynamics analysis on metabolic networks. BFODFBA specifics to maximize the production of succinate in *E. coli* has been clarified. Hybridization of BFO and DFBA has shown the ability to perform gene knockout approaches. Figure 1 displays the BFODFBA flowcharts.

The hybrid algorithm of Bacterial Foraging Optimization Algorithm and Dynamic Flux Balance Analysis (BFODFBA) can be divided into three main phases. The phases are:

1. Initialization phase
2. Fitness evaluation phase
3. New solution generation phase.

Initialization phase.

During the initialization process, a population with a matrix of 95 × 200 is created, as 95 reactions and 72 metabolites are included in the *E. coli* model. The matrix with size 95 × 200 was created by using value 0 and 1 where 0 denotes reaction that is not allowed to be knocked-out and 1 corresponds to the reaction that is allowed to be knocked-out.

There are five parameters that are required to be initialized before the implementation of the algorithm, which are *Ne* (number of elimination-dispersal events), *Nr* (number of reproduction steps), *Nc* (number of chemotactic steps), *Np* (bacteria population size) and *Ns* (number of swim steps). These parameters are set to default value according to the rosenbrock bpo.

The *E. coli* bacterium has a regulation mechanism that helps it to hunt for food and attempt to prevent toxic substances. For example, it swims away from alkaline and acidic conditions and into more neutral ones.

Fitness evaluation phase.

Dynamic Flux Balance Analysis (DFBA) is a statistical technique used to analyze the movement of metabolites across the metabolic network. Thus, the DFBA Static

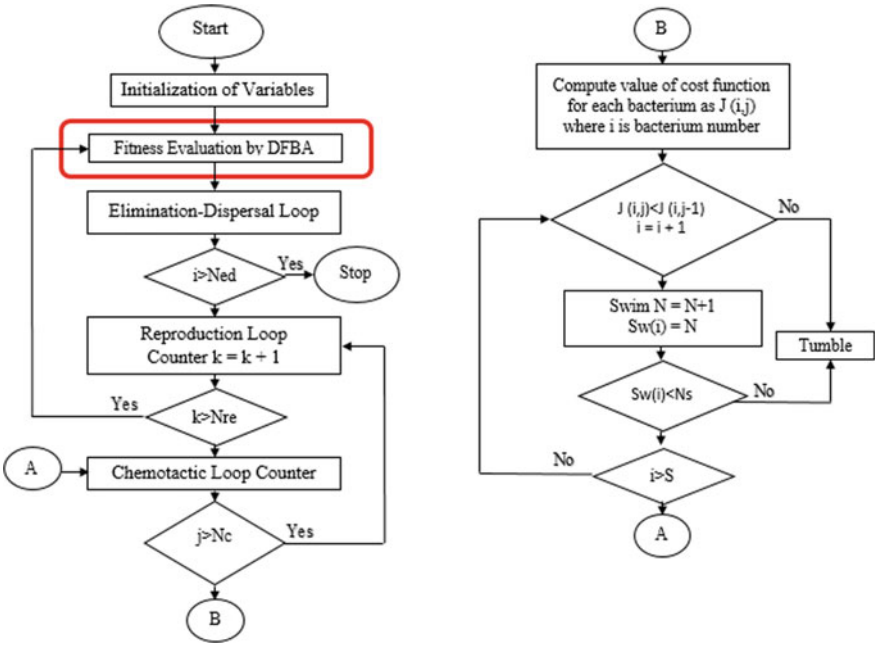


Fig. 1 Flowchart of BFODFBA. The red box represents the modified parts

Optimization Technique is used to determine the fitness of the chosen knockout reactions. In the absence of non-linear constraints, including fluxes, the optimization problem has been reduced to a linear programming problem. Then to obtain the flux distribution at a given time moment, linear programming has been replicated over the course of the batch.

New solution generation phase.

In BFO, a bacterium will undergo three main processes, which are the chemotaxis, reproduction and elimination-dispersal. Due to unexpected changes in the environment or an attack, chemotactic progress can be destroyed, and a group of bacteria may travel to some other location or another location in the swarm of concern. This is the event of elimination-dispersal in a real bacterial population where all bacteria in an area are destroyed or a colony is spread into a new part of the environment.

Initially, a random vector is generated and being applied on the bacterium to know its current position. Chemotaxis simulates the motion of the *E.coli* cell by floating and stumbling across flagella. Biologically, the *E. coli* bacteria can travel in two distinct directions. It can travel in the same direction for a while or it can tumble and shift between these two modes for a lifetime.

An initial fitness is calculated, *Jlast.* During each process, the fitness value, $J(i)$ is calculated for every bacterium, for example, add a cell-to-cell

attractant-repelling profile to mimic swarming activity. Within the number of swim steps set, if the value of $J(i)$ is smaller than the value of *Jlast*, the value of *Jlast* will be replaced with $J(i)$. If not, another random vector will be generated. $J(i)$ will be calculated as well.

In the reproduction step, the least healthy bacteria will ultimately die when one of the healthier bacteria is asexually divided into two bacteria, which would then be put in same location. This makes the volume of the swarm constant. The bacteria will be sorted in ascending cost. The bacteria with the highest value of J value dies and other bacteria with the best value will split. The value J will be updated. Gradual or unpredictable changes in the natural environment where a bacterium species is living can occur due to a variety of reasons, e.g. a significant local temperature increase may kill a population of bacteria that are currently in a high nutrient gradient area. The elimination-dispersal steps will occur, an individual bacterium is randomly chosen to be eliminated from the population and is substituted with a new bacterium positioned at a random different position within the optimization domain.

3 Results and Discussion

3.1 Experimental Results

All the findings presented are based on the in-silico simulation. BFODFBA is required to be verified to check whether it can generate desirable results before any laboratory experiments can be carried out. In order to show the efficiency of the method, the results collected were validated by comparing the in-silico results obtained from previous studies and the function of the predicted genes was validated according to literature studies. Extensive verification through lab experiments may be considered to be conducted after this. From the in-silico experiments, this paper indicates that there are three pairs of genes that have the highest production rate after 20 runs of BFODFBA with a total number of knockouts of 2–5. The findings for succinate obtained from BFODFBA are shown in Table 2.

Out of the results shown in Table 2, *CO2t*, *EX glc*(*e*) and *PGI* were found to increase succinate production to 10.43138 mmol g DW^{-1} h^{-1}. The growth rate is 0.173375 h^{-1}, which means *E. coli* cells remain alive though the genes were deleted. According to Charusanti [9], the removal of *pgi* genes affects the function of the metabolic network in *E. coli* model by inactivating the function of the upper glycolysis, a condition that affects the development of the minimum glucose media but does not kill the organism. The *pgi* gene catalyzes glucose isomerization of 6-phosphate to 6-phosphate fructose in the upper glycolysis. The inactivation of the glycolysis process improves the succinate processing process. *CO2t* has an important role to play in the development of *E. coli* in addition to its position as a substrate for carboxylation reactions [10].

Table 2 Knockout list for Succinate in *E.coli* core model

Number of gene knockout	Knockout List	Reactions	Succinate (mmol g $DW^{-1}hr^{-1}$)	Growth Rate (hr^{-1})
3	CO2t EX_glc(e) PGI (pgi)	co2[e] ⇔ co2[c] [e]: glc-D ⇔ [c]: g6p ⇔ f6p	10.43138	0.173375
4	EX_co2(e) EX_glc(e) GLUDy (gdhA) PGI (pgi)	[e]: co2 ⇔ [e]: glc-D ⇔ [c]: glu-L + h2o + nadp ⇔ akg + h + nadph + nh4 [c]: g6p ⇔ f6p	12.34212	0.24725
5	*AKGDH (lpdA)* *PTAr (pta)* *ACKr (ack)* *LDH_D (ldhA)* *PGI (pgi)*	*[c]: coa + nad + akg → nadh + succoa + co2* *PI + ACCOA ⇔ COA + ACTP* *ATP + GLU-L ⇔ ADP + GLU5P* *LAC-D + NAD ⇔ NADH + PYR + H* *[c]: g6p ⇔ f6p*	*13.4506*	*0.22396*

Note Bold font represents the best result

Another set of knockouts are *EX co2*(*e*), *EX glc*(*e*) and *GLUDy* (*gdhA*) where succinate production is increased to 12.34212 mmol g DW^{-1} h^{-1}. The growth rate of 0.24725 h^{-1} also showed an improvement. According to Ranganathan [11], the downregulation of *GLUDy* prevents the development of by-products such as lactate, formate and ethanol. The up-regulation for *ACALD* converts the residual acetate back to acetyl-CoA, which is then transformed to succinate.

The highest succinate production of 13.4506 mmol g DW^{-1} h^{-1} with a growth rate of 0.22396 h^{-1} is achieved for the 5-gene knockout. As mentioned above the removal of *pgi* genes contributes to higher succinate production. Li [12] have confirmed that the *lpdA* genes encode the components of *AKGDH*. *AKGDH* is an enzyme that specifies the tricarboxylic acid (TCA) cycle rate, where it catalyzes the succinyl-CoA (*SucCoA*) reaction of α-ketoglutarate (*α-KG*). *AKGDH* causes a reduced *NADH* power, thus disabling *AKGDH* and blocking the TCA cycle. The activity of citrate synthase (*CS*) in succinate production is affected after the deletion of the *lpdA* gene. On the other hand, the processes of *mdh* and *fum* are up-regulated. The deletion of *lpdA* also deactivates *pdhc* and catalyzes the reaction of pyruvate (*PYR*) to acetyl-CoA (*AcCoA*). Fluxes via *pdhc* and *AKGDH* reactions can be restricted to zero. Pyruvate will then be generated and will gradually increase the production of succinate until all glucose is used. The study by Sánchez [13] suggested that the disruption of the *adh*, *ldh*, and *ack-pta* genes from the central metabolic pathway could increase the glucose yield of succinate. Elimination of the *ACKr* and *PTAr* genes eliminates the main mechanism of acetate production. Removal of the acetate pathway helped to convert carbon molecules to acetyl-CoA. This affects the pathway of production of glyoxylate and succinate. Finally, glyoxylate is further transformed to malate, then to fumarate then finally to succinate, that has helped to limit the uptake of *NADH*. In addition, the elimination of *LDH D* genes also improves succinate production. Knocking out of the lactate pathway is capable of retaining *NADH* and carbon atoms, thereby assisting to direct the carbon to the acetyl-CoA pool. This affects the pathway of production of glyoxylate and succinate.

3.2 Comparative Analysis

Table 3 demonstrates the comparative experimental results of the proposed BFODFBA, OptKnock [2] and MOMAKnock [14] for succinate production.

Table 3 demonstrates the yield of succinate of different gene knockouts. The best succinate production rate for OptKnock and MOMAKnock are 6.21 and 5.02 mmol gDW^{-1} h^{-1}, respectively. The proposed BFODFBA shows a much-improved performance of succinate production of 10.43138 mmol gDW^{-1} h^{-1}, which is the best among the methods compared. This implies that BFODFBA approach has the advantage in maximizing the production rate of succinate in the *E. coli* model.

Table 3 Comparative analysis for Succinate in *E.coli* core model

Targeted Reaction	Succinate		
Method	OptKnock	MOMAKnock	*BFODFBA*
Production Rate	6.21	5.02	*10.43138*
Number of Knockouts	3	3	*3*
List of Knockout	PYK, ACKr, PTAr	SUCDi, PGL, URIC	*CO2t, EX_glc(e), PGI*

Note Bold font represents the best result

4 Conclusion and Future Work

In conclusion, BFODFBA algorithm demonstrated its capability of identifying genes to be removed in *E. coli*'s metabolic network to improve succinate production. In-silico results from BFODFBA algorithm concurred with in-vitro results. BFODFBA is robust in evaluation of ideal knockout genes set and this paper highly recommended results obtained from this algorithm to be tested out in laboratorial production of succinate in *E. coli*. However, BFO encounters slow convergence over complex optimization problems compared to other nature-inspired optimization methods [15]. Therefore, in the future, there are other constraint-based methods such as ROOM or improved BFO to be hybridized with in order to suggest different gene knockout strategies that might come out with more desired results. There are also a lot more models to be used other than *E. coli* such as *Clostridium acetobutylicum* for biofuels production [16]. A newer version of the dataset used in this research might also produce different outcomes.

Acknowledgements This work is a part of a project submitted to Universiti Teknologi Malaysia. We would like to thank the Malaysian Ministry of Higher Education and Universiti Teknologi Malaysia for their support via the Fundamental Research Grant Scheme (grant number: R. J130000.7828.4F886) and Malaysia's Research Star Award (FRGS-MRSA) (no grant: R/FRGS/A0800/01655A/003/2020/00720). This work is also supported by Universiti Malaysia Kelantan via the Postdoctoral (Research) Scheme.

References

1. Tang PW, Choon YW, Mohamad MS, Deris S, Napis S (2015) Optimising the production of succinate and lactate in Escherichia coli using a hybrid of artificial bee colony algorithm and minimisation of metabolic adjustment. J Biosci Bioeng 119(3):363–368
2. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84 (6):647–657
3. Pharkya P, Maranas CD (2006) An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng 8(1):1–13

4. Tepper N, Shlomi T (2010) Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics 26(4):536–543
5. Passino KM (2002) Biomimicry of bacterial foraging for distributed optimization and control. IEEE Control Syst Mag 22(3):52–67
6. Chen H, Zhang Q, Luo J, Xu Y, Zhang X (2020) An enhanced Bacterial Foraging Optimization and its application for training kernel extreme learning machine. Appl Soft Comput 86:105884
7. Zareh S, Seyedjavadi HH, Erfani H (2012) Grid scheduling using cooperative BFO algorithm. Am J Sci Res 62:78–87
8. Mahadevan R, Edwards JS, Doyle FJ III (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83(3):1331–1340
9. Charusanti P, Conrad TM, Knight EM, Venkataraman K, Fong NL, Xie B, Gao Y, Palsson BØ (2010) Genetic basis of growth adaptation of Escherichia coli after deletion of pgi, a major metabolic gene. PLoS Genet 6(11):e1001186
10. Kozliak EI, Fuchs JA, Guilloton MB, Anderson PM (1995) Role of bicarbonate/CO2 in the inhibition of Escherichia coli growth by cyanate. J Bacteriol 177(11):3213–3219
11. Ranganathan S, Suthers PF, Maranas CD (2010) OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6(4):e1000744
12. Li M, Ho PY, Yao S, Shimizu K (2006) Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by 13C-labeling experiments. J Biotechnol 122(2):254–266
13. Sánchez AM, Bennett GN, San KY (2005) Novel pathway engineering design of the anaerobic central metabolic pathway in Escherichia coli to increase succinate yield and productivity. Metab Eng 7(3):229–239
14. Ren S, Zeng B, Qian X (2013) Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints. In: BMC bioinformatics, vol 14, no S2. BioMed Central, p S17
15. Chen H, Zhu Y, Hu K (2011) Adaptive bacterial foraging optimization. In: Abstract and applied analysis, vol 2011. Hindawi
16. Yoo M, Soucaille P (2020) Trends in systems biology for the analysis and engineering of Clostridium acetobutylicum metabolism. Trends Microbiol 28(2):118–140

Analysis of Outpatient Visit Pattern and Waiting Time Prediction for Selected Public Health Clinics in Selangor

Suhaila Zainudin and Dzulhusni Anjang Ab. Rahman

Abstract Public clinics are the preferred choice for health treatment worldwide. This preference results in many outpatients seeking remedies at public clinics and is the leading factor in patients' long waiting time. The definition for waiting time is when a patient has to remain at the clinic for treatment. The waiting time starts with the registration process until consultation with the physician. Lengthy waiting time is the leading cause of patient dissatisfaction. The discontent is especially so when there is no convenient waiting facility. The objective of this study is to analyze outpatient visits patterns to public outpatient clinics. The recognized patterns will form the basis to recommend which public outpatient clinics are the best to visit at a specific time to minimize waiting time. The waiting time is also predicted based on the arrival time. Fifteen public outpatient clinics in Selangor that use an electronic medical record system provided this study's data. The data analysis shows a correlation between patient waiting time, day, the month of visit, patient age, and consultation time. Past research shows that classification of machine learning methods can predict the waiting time. This study has demonstrated that Linear Discriminant Analysis creates the best classification model for Puchong and Batu 9 Cheras datasets. Support Vector Machine (SVM) is the best classifier for the Anika dataset.

Keywords Waiting time · Public outpatient clinic · Arrival time · SVM · LDA

S. Zainudin (✉)
Center for Artificial Intelligence Technology, Fakulti Teknologi Dan Sains Maklumat, Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor, Malaysia
e-mail: suhaila.zainudin@ukm.edu.my

D. A. Ab. Rahman
Bahagian Pengurusan Maklumat, Kementerian Pendidikan Malaysia, Aras 3 and 4 Blok E11, Kompleks E, Pusat Pentadbiran Kerajaan Persekutuan, 62604 Putrajaya, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_48

1 Introduction

The life expectancy statistics of the Malaysian population increases every year. Individuals born in 2018 may live 2.8 years longer until they reach 75 years compared to 72.2 years for individuals born in 2000 [1]. The longer life expectancy in the nation is related to the increased demand for health services in public outpatient clinics. Outpatients clinics in public hospitals are often overcrowded. Overcrowding significantly contributes to poor patient experience with physician services [2]. This paper has five sections starting with Sect. 1 for Introduction and Background, Sect. 2 for Problem Statement, Sect. 3 for Research Methodology, Sect. 4 for Data Analysis, and the last section is for Conclusion.

1.1 Background

Reference [3] defined the waiting time as when the patient has to wait from registration until consultation with the medical officer. The global issue of long waiting time in hospitals has spurred research in this domain. Reference [4] proposed a Patient Treatment Time Prediction or PTTP algorithm predict the required time for each treatment task for a patient and subsequently developed a Hospital Queuing-Recommendation (HQR) system. An improved Random Forest (RF) algorithm is the basis for PTTP. The PTTP and HQR have undergone extensive experiments and shown high precision and performance.

Easing congestion in emergency departments is an active research area [5]. Research in this area works on various aspects, such as the length of patients' stay [6]. Closer to home, the Ministry of Health Malaysia has implemented multiple initiatives such as the Teleprimary Care System (TPC) since 2004. TPC is an electronic system used in healthcare to facilitate the workflow in a healthcare organization and improve patient care quality [7]. The system is an electronic medical record system that records outpatients' activities from check-in to service completion. TPC is the primary source of data for the work reported in this paper.

Reference [8] investigated several machine learning algorithms such as neural network, random forest, support vector machine, elastic net, multivariate adaptive regression splines, k-nearest neighbours and others to find the most accurate methods to predict waiting times. They concluded that the elastic net model is the best model to predict waiting time or delay time.

Reference [9] compared several machine learning algorithms, including random forests, elastic net, gradient boosting algorithm, support vector machine and multiple linear regression, to find the most accurate model for predicting waiting time in pediatric ophthalmology outpatient clinics. Their work concluded that supervised machine learning models could accurately predict wait time and identify the most significant contributing factors.

2 Problem Statement

Public outpatient clinics are the first choice for health treatment in a community. The preference for outpatient clinics causes long waiting periods and subsequently leads to overcrowding. A long waiting time with no available comfortable waiting space is a significant cause of public outpatient clinics' complaints. Uncertainty of expected waiting time for treatment is one of the possible reasons for dissatisfaction in outpatients. Sickly outpatients who are waiting their turn also have the potential to spread the disease to other outpatients. If the clinic can estimate the waiting period during the initial registration to the patient, the outpatient can wait at various locations or return home whilst waiting for their turn. However, possible waiting times are difficult to be predicted as the waiting period depends on the type of treatment for each outpatient, the procedures to be followed for individual ailments and the doctor's consultation or diagnosis style [10].

The waiting time is when the patient has to wait from registration or check-in until consultation with the medical officer [3]. In Malaysia, the average waiting time is between 1 to 2 hours [10]. According to [11], long waiting time is the main reason for patient dissatisfaction compared to other reasons such as clinic facilities and staff attitude. Informing the estimated waiting time to the patient earlier can minimize the patient's dissatisfaction. However, it is difficult to predict the patients' waiting time since the waiting time depends on the type of treatment, procedures of the previous patient and the doctor consultation or diagnosis style [4].

The patient's historical record of visits to the clinics is a potential source of data containing hidden past visits. Based on the TPC data set, there are different peak hours between various public outpatient clinics in the same district. Analyzing these records may unearth hidden patterns unique to the patient's visits. The results from these analyses are potentially helpful. Possible uses are to suggest the time to visit a particular clinic since this will minimize waiting time. Another benefit is to predict the waiting time based on patient arrival time.

3 Research Methodology

Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology formed this study's basis (Fig. 1). The study started with a business understanding phase, data understanding phase, and data preparation phase in line with the CRISP-DM method. The following two steps are the attribute selection phase and the modelling phase. The data source is obtained from the TPC records from 15 outpatient clinics in Selangor from January to December 2018. The raw data contained more than a million lines of records (1,320,459 records) in Microsoft Excel format.

The data understanding phase consists of several activities, namely data acquisition, descriptive analysis and data visualization. Then, the data goes through the data preparation phase, which aims to produce the final dataset to be analyzed by

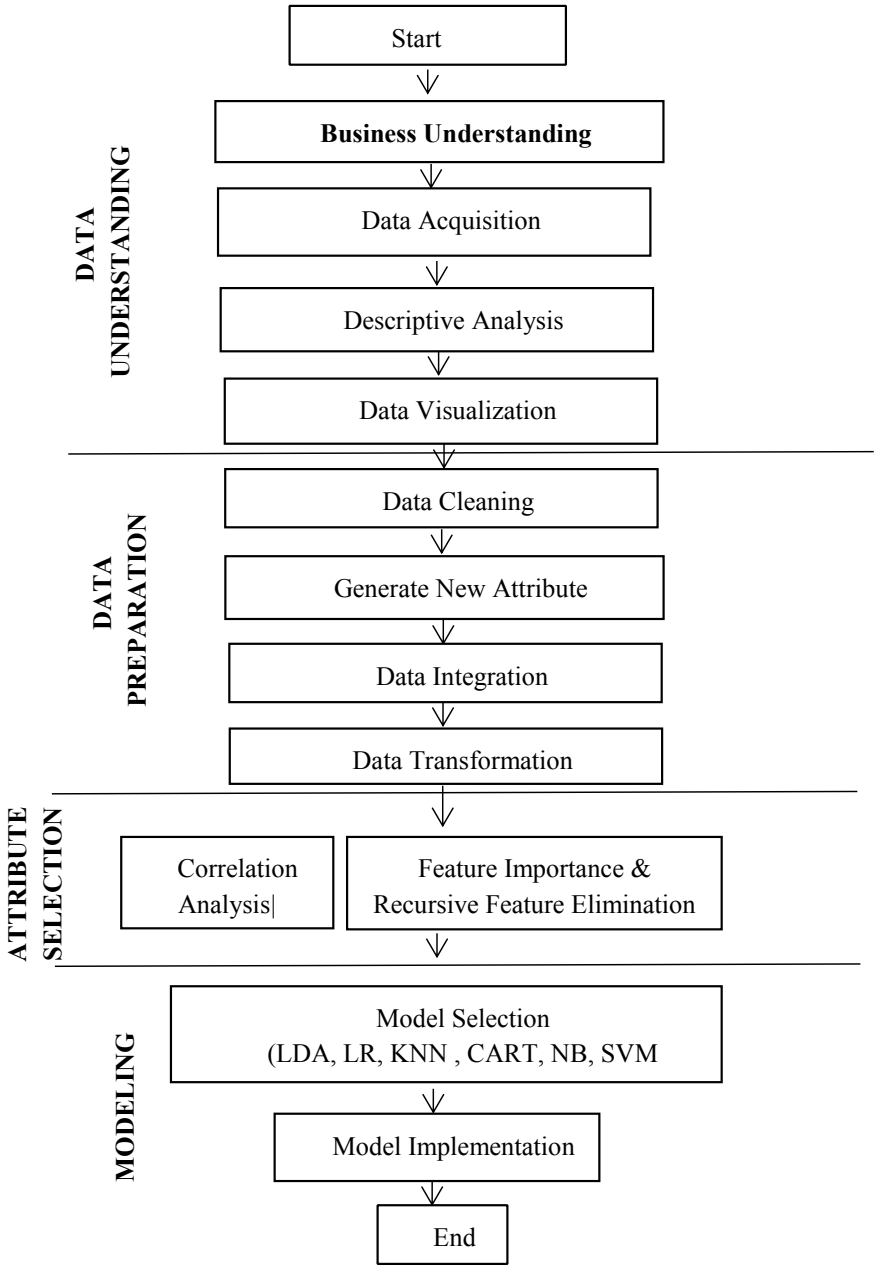


Fig. 1 Research methodology

Table 1 Discretised wait time values

Discretised value	Value range
60P	0–53 min
120P	54–106 min
160P	107–160 min

the machine learning algorithms. Activities in this phase are data cleaning, generating new attributes, integrating data sets, and data transformation into a format that is compatible with machine learning algorithms. The first task in data preparation is data cleaning to remove attributes that are not relevant to the study. For example, PatientName, PositionConsultation, Module dan PerformByConsultation. The study eliminated any records that contain the attribute TimeWaiting with negative values. Since the public outpatient clinics start their operation at 7.30 am, the study removed any records that contain the attribute ArrivalTime with value before 7.30. Although outpatient clinics are open on weekends by appointment, this study is only focusing on weekdays. Therefore, the study did not consider weekend records.

The next activity is generating new attributes. The original data set has no available detail for medical officer consultation time length. To generate the new attribute ConsultationLength, this formula is; attribute ConsultationTime minus attribute ArrivalTime. The study separated the ArrivalTime attribute into two parts: hour and minute, without deleting the original feature. The new attributes are named ArrivalHour and ArrivalMinute. Then, the data set is integrated with other data set containing the day of the week, day of the month and month.

The values of CreatedDate and ReportDate are transformed from date format to number format using the Microsoft Excel datevalue function. The ArrivalTime, ConsultationTime and ConsultationEndTime attributes are converted from time to number format using the Microsoft Excel timevalue function. The TimeWaiting attribute is discretized form continuous integer to three categorical value, namely 60P, 120P and 160P (Table 1).

In Attribute Selection, the study used correlation analysis with the prepared data to find attributes that correlate with TimeWaiting. Further modelling will use these most correlated attributes. The Python functions for attribute selection, such as Feature Importance (FI) and Recursive Feature Elimination (RFE) functions, were utilized for attribute selection. The final attributes selected for the study is the combination of outputs from the FI and RFE functions (Table 2). The choice of these attributes aims to reduce the experiment's processing time and ensure better accuracy.

3.1 Data Modelling

For Data Modelling, the TimeWaiting attribute is discretized into three types of categorical values of "60P", "120P", and "160P". These categorical values are class

Table 2 Selected attributes

No	Feature Importance
1	ArrivalTime
2	ConsultationTime
3	ConsultationEndTime
4	AgeFinal
5	Month
6	DayOfWeek

attributes. The data set is divided into two parts, training data sets that contain seventy per cent (70%) of the data and the remaining thirty per cent (30%) of the data set as test data sets. Modelling uses the tenfold cross-validation approach to evaluate the accuracy using training data sets. The training data set is used with classifications machine learning algorithms. Later, the test data will test the performance of the best models derived from the experiments. Descriptive analysis showed that the data set is skewed to the right with a small number of records. This skewed data record is an example of noisy data. Therefore, the study removed TimeWaiting records with a value greater than 160 (representing the noisy data).

The classification algorithms are LDA (Linear Discriminant Analysis), LR (Logistic Regression), KNN (Nearest Neighbors), CART (Decision Tree), NB (Gaussian Naive Bayes) and SVM (Support Vector Machine). The modelling phase used the top three most correlated data sets: Puchong clinic, Anika clinic, and Batu 9 Cheras clinic. The three classification algorithm's model's accuracy with the highest score for each clinic is in Tables 3, 4 and 5.

Table 3 Results of classification Puchong dataset

No	Algorithm	Accuracy (%)
1	LDA	79.21
2	CART	15.99
3	NB	1.79

Table 4 Results of classification on Anika dataset

No	Algorithm	Accuracy (%)
1	SVM	95.84
2	KNN	95.16
3	LDA	93.67

Table 5 Results of classification on Batu 9 Cheras dataset

No	Algorithm	Accuracy (%)
1	LDA	99.02
2	CART	98.62
3	LR	79.06

For the Puchong dataset, the LDA algorithm gives the best results with a model accuracy of 79.21%. The second-best model is CART, with a model accuracy of 15.99%. The NB model scored a very low accuracy of less than 2%. For the Anika dataset, the SVM model gives the best result of 95.84%, followed by KNN with a score of 95.16% and lastly, LDA with a score of 93.67%. The LDA model scored the best classification accuracy with a value of 99.02% for the Batu 9 Cheras data set. The CART model also provides good accuracy of 98.62%. The model produced by the LR scored lower accuracy of 79.06%.

4 Data Analysis

As described previously, the study separated the data set into 70% for the training data set, and the remaining 30% is for testing. The study uses the best model produced during the modelling phase with the test data. The results for testing are described further in this section.

4.1 Evaluation Using Test Set

Tables 3, 4, and 5 show the LDA as the best performing classifier for Puchong and Batu 9 Cheras training data set, and SVM is the best performing classifier for Anika. Therefore, the study chose LDA as the best classifier for evaluation using the Puchong and Batu 9 Cheras clinics' test data. Meanwhile, the study also chose SVM as the best classifier for the Anika test data set.

Puchong Test Data Set

The LDA classification model scored an accuracy of 98.53% model on the test data (Table 6). Based on the confusion matrix in Table 6, LDA can correctly classify class for 60P. The class 60P represents TimeWaiting with a waiting time interval from 0 to 53 min. Table 7 shows that this model has the lowest performance in predicting the class of 160P. The possible reasons are that the Puchong test data set contained a small fraction of sample data with a class value of 160P compared to class 60P.

Table 6 Confusion matrix for Puchong test data set (LDA)

	120P	160P	60P
120P	3327	55	66
160P	17	794	0
60P	21	0	6526

Table 7 Accuracy, Sensitivity, F1 score and Support for the LDA model for Puchong data set

	Accuracy	Sensitivity	F1 score	Support
120P	0.99	0.96	0.98	3448
160P	0.94	0.98	0.96	811
60P	0.99	1.00	0.99	6547
Average/ Total	0.99	0.99	0.99	10,806

Table 8 Confusion matrix Anika test data set (SVM)

	120P	160P	60P
120P	17,117	0	7
160P	429	0	1
60P	839	0	12,490

Table 9 Accuracy, Sensitivity, F1 score and support for the SVM model for Anika data set

	Accuracy	Sensitivity	F1 score	Support
120P	0.93	1.00	0.96	17,124
160P	0	0	0	430
60P	1.00	0.94	0.97	13,329
Average/ Total	0.95	0.96	0.95	30,883

Anika Test Data Set

The SVM model produced an accuracy of 95.86% on the Anika Test data set. The confusion matrix in Table 8 shows that the SVM model can classify class attributes with a record value of 120P for TimeWaiting with an estimated waiting time interval from 19 to 89 min very well. Based on Table 9, the test data set does not have comprehensive sample data for class 160P. Hence the model is not able to test that particular class.

Batu 9 Cheras Test Data Set

The LDA model performed satisfactorily on the Batu Cheras test data set with an accuracy of 99.07%. The model is very consistent in classifying all test data for all classes Table 10. As shown in Table 11, the LDA classification algorithm's model is the best compared to Puchong and Anika's models. Apart from the high percentage of accuracy, this model can also predict all classes consistently.

From the experimental results, we conclude that although each health clinic has similar data features, each clinic has its distinct characteristic and behaviour. Hence, other clinics cannot utilize the models generated from a different clinic. This behaviour is likely due to different organization culture among clinics. The developed machine learning model is unique only to the specific clinic and does not apply to other clinics. The discretization of the TimeWaiting values into three

Table 10 Confusion matrix for Batu 9 Cheras test data set (LDA)

	120P	160P	60P
120P	5506	51	111
160P	1	1340	0
60P	8	0	11,420

Table 11 Accuracy, Sensitivity, F1 score and Support for the LDA model for Batu 9 Cheras data set

	Accuracy	Sensitivity	F1 score	Support
120P	1.00	0.97	0.98	5668
160P	0.96	1.00	0.98	1341
60P	0.99	1.00	0.99	11,428
Average/ Total	0.99	0.99	0.99	18,437

categories produced better classification results. However, there is a downside when using discretized wait time since the estimated wait time for the TimeWaiting attribute can be only estimated in a limited range (Table 1).

The Anika clinic test data set is unbalanced in term of its classes, specifically by the number of records in the 160P class. Model accuracy test on TimeWaiting attribute with the class of 160P was not possible due to a lack of records in that class. The study gained further insights from data analysis that may help avoid unnecessarily long waiting time for outpatients. Outpatients should come early to the clinics in densely-populated areas as these clinics will eventually be over-crowded by the public after 7.45 am. Based on the correlation analysis, the Puchong clinic is senior citizen-friendly since senior citizens' waiting time is shorter. All of the clinics are busiest on Mondays and Thursdays. The most active months are January and July, except for the Petaling district, where most patients frequented the clinics in July, August, October, November and December. A recommendation to avoid a long wait is that patients should arrive in the early morning before 7.30 am or in the afternoon around 4:00 pm.

5 Conclusion

This study can be used as initial input for the Ministry of Health to develop an online waiting time prediction system, especially for Puchong clinic, Anika clinic and Batu 9 Cheras clinic. Outpatient visit patterns are unique based on arrival time, day and month. The total number of medical officer consultations reaches its maximum of only one to two hours each day. An investigation into this area may identify the reason and make further corrective actions on working hours and scheduling of medical officer duty roster. For every district, there are less visited clinics. The less-visited clinics can loan or share their resources such as unused

equipment or workforce owned to busier clinics. Several tasks can continue the current study; for example, to record all start and end time for each procedure that a patient has undergone.

Acknowledgements This work is supported by UKM Research University Grant (GUP-2020-089).

References

1. Department of Statistics Malaysia (2018) PRESS RELEASE. SHORT LIFE SCHEDULE, MALAYSIA, 2016–2018. Department of Statistics. Putrajaya
2. Bao Y, Fan G, Zou D, Wang T, Xue D (2017) Patient experience with outpatient encounters at public hospitals in Shanghai: Examining different aspects of physician services and implications of overcrowding. PLoS One. 12(2):e0171684. Published 2017 Feb 16. https://doi.org/10.1371/journal.pone.0171684
3. Belayneh M, Woldie M, Berhanu N, Tamiru M (2017) Global journal of medicine and public health 1 www The determinants of patient waiting time in the general outpatient department of Debre Markos and Felege Hiwot hospitals in Amhara regional state, North West, Ethiopia. Gjmedph.Com 6(5). Retrieved from www.gjmedph.com
4. Chen J, Li K, Tang Z, Bilal K, Li K (2016) A parallel patient treatment time prediction algorithm and its applications in hospital queuing-recommendation in a big data environment. IEEE Access. https://doi.org/10.1109/ACCESS.2016.2558199
5. Morley C, Unwin M, Peterson GM, Stankovich J, Kinsman L (2018) Emergency department crowding: a systematic review of causes, consequences and solutions. PLoS One 13(8): e0203316. Published 2018 Aug 30. https://doi.org/10.1371/journal.pone.0203316
6. Mentzoni I, Bogstrand ST, Faiz KW (2019) Emergency department crowding and length of stay before and after an increased catchment area. BMC Health Serv Res 19(1):506. Published 2019 Jul 22. https://doi.org/10.1186/s12913-019-4342-4
7. Ho AC (2013) Acceptance and utilization of teleprimary care in Sarawak. PhD thesis, Universiti Malaysia Sarawak [cited 1 March 2021]. Available from https://ir.unimas.my/id/eprint/15439/
8. Curtis C, Liu C, Bollerman TJ, Pianykh OS (2017) Machine learning for predicting patient wait times and appointment delays. J Am Coll Radiol 15(9):1310–1316. https://doi.org/10.1016/j.jacr.2017.08.021. Epub 2017 Oct 24. PMID: 29079248
9. Lin WC, Goldstein IH, Hribar MR, Sanders DS, Chiang MF (2020) Predicting wait times in pediatric ophthalmology outpatient clinic using machine learning. In: AMIA ... Annual Symposium proceedings. AMIA Symposium, 2019, pp 1121–1128
10. Ahmad BA, Khairatul K, Farnaza A (2017) An assessment of patient waiting and consultation time in a primary healthcare clinic. Malays Fam Phys Official J Acad Fam Phys Malays 12 (1):14
11. Pillay MS, Mohd Ghazali RJ, Abd Manaf NH, Abdullah AHA, Abu Bakar A, Salikin F, Umapathy M et al (2011) Hospital waiting time: the forgotten premise of healthcare service delivery? Int J Healthc Qual Assur 24(7):506–522. https://doi.org/10.1108/09526861111160553

Case Study: The Implications of Knowledge Management Tools on the Process of Overcoming COVID-19

Fatima Almatrooshi, Sumayya Alhammadi, Said A. Salloum, and Khaled Shaalan

Abstract The main aim of this paper is to identify the most important issues and problems that we faced in the UAE in the case of coronavirus pandemic and try to come up with solutions that will enhance the overall response of organizations and individuals to this crisis. Moreover, we want to investigate the importance of knowledge management in spreading information regarding COVID-19 and how knowledge management has been used to manage this pandemic as it continues to manifest itself in several areas and different capacities. In addition, this paper will highlight the significance of effective management of knowledge and information with regards to COVID-19, and the extent to which the knowledge that has been shared widely through different platforms can help in resolving the pandemic. These findings will also enable identification of the efficacy of knowledge management and its achievements in dealing with previous pandemics in a manner that gauges the progressive stages of how it was implemented. The quantitative and qualitative data that was gathered through a survey helped us to identify the problems and uncover hidden issues faced by individuals and organizations. The recommended solution we thought the best in terms of cost and efficiency is deploying crisis management in a professional way using knowledge management tools in organizations in public society. We explained the full process of the solution in this paper and the plan of execution and implementation.

Keywords Knowledge management · UAE · COVID 19 pandemic

F. Almatrooshi · S. Alhammadi · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_49

1 Introduction

The year 2020 has been a challenging year with the COVID-19 crisis that affected individuals and organizations in many different ways [1–3]. The effect is not necessarily negative, as it is always associated with the term crisis, but it can be positive as well. COVID-19 helped individuals, communities and even organizations realize the lack of crisis management culture on many levels. As mentioned in [4] study, there are some hidden opportunities in a crisis like COVID-19, one being able to see the gaps and build a more alert and understanding community to these diseases or any other crisis. COVID-19 belongs to a family of coronaviruses that come from animals. This virus made its way into human beings, with symptoms such as common cold and severe acute respiratory syndrome. The COVID-19 is formidable to reckon with because it has a high rate of infection, and it has subjected the entire humanity to the grave vulnerability of its attack. More than 3 million people have been infected at the time of writing this, and more than 1 million people have died from the virus globally. Two months after the first case was confirmed in Wuhan, the World Health Organization termed the COVID-19 as a Public Health Emergency of International concern.

Humanity dealt with many similar cases, such as EBOLA, SARS and HIV where a disease resulted in lives losses and financial damages. However, every time it occurs, the affected region seems to struggle at dealing with it. It is understandable that each case has its own causes and its own results, but that doesn't change the fact that government defensive procedures and organizational cautious cannot benefit from past cases. And this is where knowledge based organizations benefits from knowledge management systems and technologies. When dealing with diseases like COVID-19, EBOLA or SARS, Knowledge is a critical factor of succeeding with the process of minimizing the damages done by them. An organization with a good Knowledge Management infrastructure will surly have the advantages as it was noticeable that they were more flexible and faced less issues with the sudden changes that happened with the pandemic of COVID-19. Companies and several stakeholders around the world have sought after knowledge management solutions to enable them to deal with a particular technology, processes undergone in the initiative to utilize this technology, as well as the structure that governs this initiative in the hope that it will enable them to utilize a given innovation. According to Albris et al. [5], knowledge management in various media, including social media, the internet, and TV, has been stepped up with an effort to deliver real-time and accurate information in its bid to curtail further infections in the current global pandemic of the COVID-19. There are people around the world, including in the UAE and several media houses, who have played a significant role in managing information in the right way. These individuals effectively and efficiently and digital knowledge management for its capacity to restrict the exceptional infection, and also to remedy the prevalence of this pandemic. The challenge for an organization is to develop a Knowledge Management System (KMS) that can easily adapt to change in dealing with uncertainties [6]. Many people in such crisis

choose to communicate through social media and knowledge sharing tools as reported in [7]. The authors stated how the governments can benefit from these social platforms in the process of damage analysis based on what is shared from the public communities.

1.1 Domain Characteristics

The medium we chose to understand the problem more and investigate for solutions is the society and organizational communities' domain. The main issues and problems faced by this domain were discovered and analyzed through a survey that we conducted over a period of 1 week. This survey targeted different people from different social and career backgrounds. After identifying the major problems we analyzed them and matched them with recommended solutions based literature reviews and research. Participants in this study were evenly distributed regarding age. This is a vital factor in this study to understand how different age groups acquire information and how the pandemic has affected them, Fig. 1. Moreover, Most of the participants in this study have a bachelor's degree. Others have masters, higher degrees and Ph.D., as shown in Fig. 3. This is important as it helps understand how these individuals perceive the impact of the virus, and their level of education also matters regarding how they interact with knowledge management tools such as social media to obtain information about COVID-19. The number of the participants, more specifically, 82 participants, do not work. 12 participants work in the private sector, 8 work in semi-government companies, and 67 work in the government. In other words we got an even distribution of working and non-working individuals, with 85 unemployed and 87 employed in different sectors, as shown in Fig. 2.

Our study does also benefit from recent research that has been done which has a more general and worldwide view of the impacts of the pandemic.

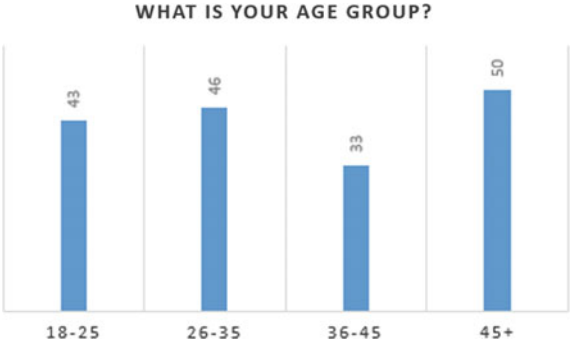


Fig. 1 Survey participant's age groups

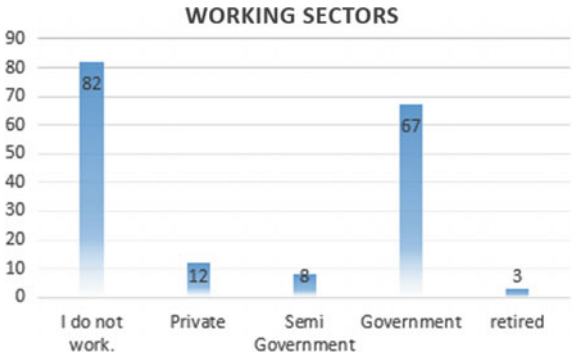


Fig. 2 Participant's career status

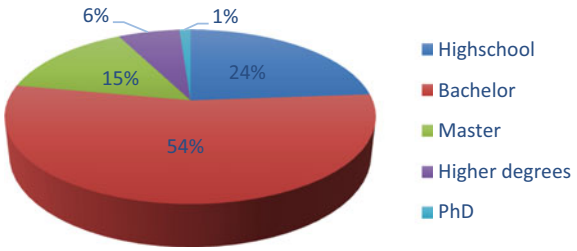


Fig. 3 Survey participant's educational level

1.2 Problem Identification and Analysis

We conducted a survey over a period of 1 week that targeted different individuals from different career backgrounds. The survey was in both Arabic and English language to maximize the number of people participating and to make sure that individuals with different educational status can also share their thoughts on the topic. Moreover, the nature of our public society always prefers their native language over others, and that shows in the cumulative number of participants of the survey, as the Arabic version got 135 participants, while 38 for the English version. The survey questions were relative to the topic and easy to understand. The aim of this survey is to investigate and understand different individual's points of view about the pandemic and how it affected their lives in different aspects. Moreover, we want to understand the most important issues faced during this crisis and the reason behind the struggle with some organizational process such as finance and Information technology. In the following section we show some of the important results or our survey and how it helped in the problem and solution identification process.

2 Methods

This section mainly involved the collection of relevant data, which was used in the analysis and conclusions section. This section will discuss the design of the study, the strategies for data collection, as well as sampling method used to select the most suitable participants for this study. The research strategy used for this study is quantitative research. Quantitative research entails the collection of numerical data, which is analyzed using methods such as tables, graphs, and statistical analysis. In this case, quantitative data was collected and analyzed to come up with accurate conclusions that may be used to inform policymaking with regard to knowledge management tools for managing the COVID-19 pandemic. The approach of research used in this case was a survey. An online survey was created, and the link was sent to the participants of this research on WhatsApp, as well as through email. The participants that confirmed they would take part in the research were encouraged to be prompt to ensure that they answer the questions asked well without any hurry. The research was conducted in two languages that are in English and in Arabic. This was considered by us to ensure that the study covered a larger scope of individuals to come up with data that can be trusted. This was also a way to ensure that the information captured is valid. Sampling entails selecting a small representative population of a whole and using information acquired from the representative population to make conclusions about the whole population. The sampling technique used in this research is simple random sampling. This technique consists of randomly selecting individuals from a population. In this technique, all individuals have an equal chance of being selected.

2.1 Survey Highlights

The participants were given a set of questions that mainly aim to identify their view of the pandemic and what are the gaps they feel exist in currently deployed procedure of dealing with it. For example, participants answered a question regarding what social media tool they used the most during the pandemic and to what extent they trust this source of knowledge. Figure 4 reveals the knowledge management tools used to disseminate information about COVID-19 to participants. Most of the participants stated that they got information regarding viruses on Instagram, Twitter, WhatsApp, newspapers, and Snapchat. These platforms have been used effectively by the government of UAE to disseminate information regarding the statistics about COVID-19, that is, the number of new infections, the number of deaths from the virus, as well as the number of recoveries. Additionally, these platforms have been used to disseminate information regarding the measures that should be taken to curb the spread of the virus, such as washing hands constantly, sanitization, wearing masking public places, and eating healthy foods and a lot of vitamins to boost one's immune system. These platforms have been very effective

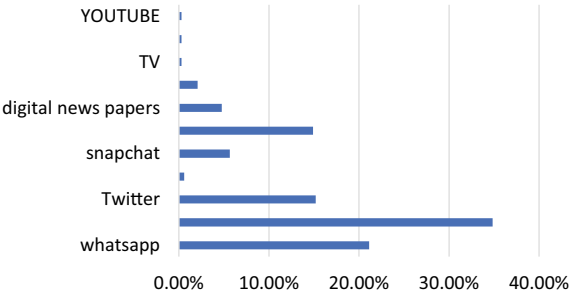


Fig. 4 Most used tools as a source of knowledge

in managing the COVID-19, as well as the impact of this virus on various sectors, such as traveling, tourism, education, and the fashion industry.

However, it was important to know the level of trust these individuals have for these resources, as many of these tools could be used to spread rumors and encourage breaking rules and restrictions. Figure 5 shows that the majority of participants have trust percentage over the 50%, which motivated us to research more about the effects of rumors spreading and misjudgment of the situations caused by trusting social media tools. However, a good percentage of the participants do not trust news and information shared on these platforms. One of the reasons why they may not trust these sources is because of a lot of fake news coming from these sources. Additionally, there is an emergence of conspiracy news being spread through these platforms, and the public has become skeptical of the information they see on these platforms. Therefore, they do not completely trust the information shared, and the majority prefer verified sources such as TV news, as well as a digital newspaper. Others prefer to verify this information by double-checking information shared by the Ministry of Health and other global organizations such as WHO.

Another question were there to identify if participants used new tools during the pandemic, what type of tools were they and how well these tools were. Additionally, the overall performance of individuals was questioned, and if the

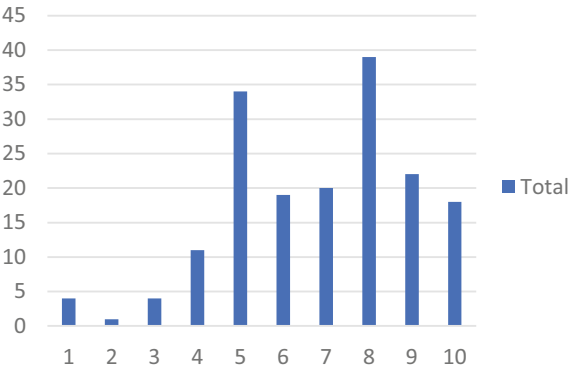


Fig. 5 Level of trust in SM knowledge sources

current tools and mechanisms used are suitable for the future. We also gave the participants a chance to express their opinion about the overall situations and what are their insights of the pandemic and what they suggest for future improvements. The results of these questions as shown in Figs. 6 and 7 indicates couple of concerns. The pandemic forced some individuals and organizations to use new communication and knowledge sharing and storing tools, which in return impacted how individuals perform. Most of the participants learned about a new knowledge management tool during this period. Some of these new tools are used for sharing knowledge, social media, data storage, and management of knowledge in enterprises. The results shows that almost 118 participant thought that they did not improve or uncertain of how these tool helped them, while 54 thought that they actually improved. These numbers doesn't necessarily judge the situation, because these insights usually depend on many factors and not only the tool used, however these numbers shows that some people are viewing this situation negatively and that it can be improved. See Table 1.

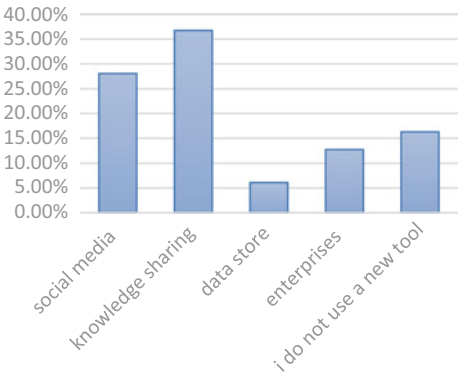


Fig. 6 Type of tools used during the pandemic

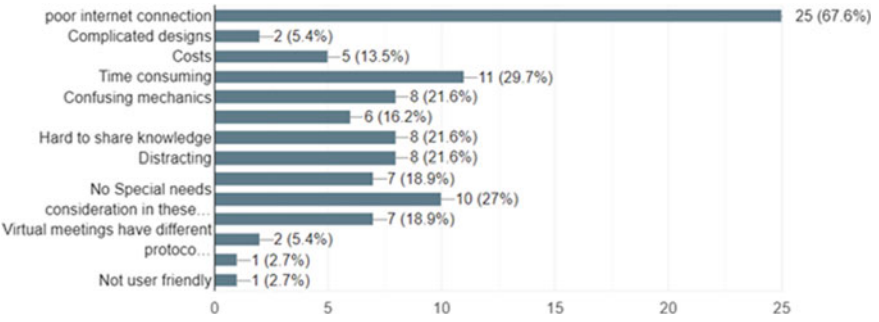


Fig. 7 Issues faced while using technology tools

Table 1 Number of participant of feel that their performance improved during the pandemic and because of tools usage

Improved performance?	Number of participants
Maybe	68
No	50
Yes	54

In addition, it was very important to know what are the issues that people faced while using tools and platforms in the COVID-19 situation. As shown in Fig. 7 almost half of the participants complained about the internet connection and how using tools that relay on internet affected their overall experience and performance. Time consumption is the second most reported issue that people faced, and we think it is mainly related to the first issue, as the continuos disconnection, lag and slow loading screens takes a valuable seconds from the users time. Moreover, time consumption could happen when the user is not familiar with the platform or doesn't have the proper knowledge to use it. Other interesting issues reported by some individuals is that organizational visual meetings have different protocols that are not always satisfied by the technology they use, and this issue motivates the research on either the company or organizations and studying the crisis situation correctly or not, and whither they use the proper and most suitable knowledge management tools to their tasks.

After analyzing the results and discussing and evaluating the situations we concluded a number of issues that need to be looked at and focused for future improvement. These issues are: (1) adaptation and adjusting issues (2) organizational knowledge loss (3) Financial damages (4) collaboration fail (5) Knowledge resources quality. In the Future work we want to discuss each problem, why it is important and how they are affecting organizations and communities. However, the major problems that we will focus and find detailed solutions in are relative to knowledge management, which are: adaptation and adjusting issues and organizational knowledge loss.

3 Conclusion

In this case study we conducted a survey that had 172 participants from different ages, professions and sectors. The survey targeted the topic of COVID-19, and how a crisis like this could affect the implications of knowledge management. After analyzing the survey results we concluded that the most concerned sectors in such situation is the educational sector. Moreover, the implications of knowledge management in such situations takes a different route from the traditional methods and ways since all the human interactions and knowledge exchanging is restricted by the new protocols such as virtual meetings and learning. Finally the need for

crisis management culture to exist to be available in every organization and every individual mind is essential for better interaction with future crisis to reduce damages and losses. Its importance is not only recognized by us, but also by number of people whom participated in the survey we conducted. If we were able to provide a method in which we increase the number of people who have this knowledge through different knowledge management tools and mechanisms, we will be able to deal with crisis and emergency situation better, and reduce losses and damages.

Acknowledgements This work is a part of a project submitted in fulfilment of Ph.D. Computer Science—Faculty of Engineering and Information Technology at The British University in Dubai.

References

1. Habes M, Alghizzawi M, Ali S, SalihAlnaser A, Salloum SA (2020) The relation among marketing ads, via digital media and mitigate (COVID-19) pandemic in Jordan. Int J Adv Sci 29(7):2326–12348
2. Al-Maroof RS, Salloum SA, Hassanien AE, Shaalan K (2020) Fear from COVID-19 and technology adoption: the impact of Google meet during coronavirus pandemic. Interact Learn Environ
3. Akour I, Alshurideh M, Al Kurdi B, Al Ali A, Salloum S (2021) Using machine learning algorithms to predict people's intention to use mobile learning platforms during the COVID-19 pandemic: Machine learning approach. JMIR Med Educ 7(1)
4. Buheji M, Ahmed D (2020) Foresight of coronavirus (COVID-19) opportunities for a Better World. Am J Econ 2020(2):97–108
5. Albris K, Lauta KC, Raju E (2020) Disaster knowledge gaps: exploring the interface between science and policy for disaster risk reduction in Europe. Int J Disaster Risk Sci 11(1):1–12
6. Dorasamy M, Raman M, Kaliannan M (2013) Knowledge management systems in support of disasters management: a two decade review. Technol Forecast Soc Change 80(9):1834–1853
7. Kaewkitipong L, Chen CC, Ractham P (2016) A community-based approach to sharing knowledge before, during, and after crisis events: a case study from Thailand. Comput Hum Behav 54(2016):653–666

Sentiment Analysis of Arabic COVID-19 Tweets

Dena Ahmed, Said A. Salloum, and Khaled Shaalan

Abstract With the COVID-19 outbreak in 2020, information about the pandemic has been exponentially increasing and spreading across various social media platforms. People across the globe have been affected in a way or another because of different aspects such as the increase in infected cases, death rate increase, financial difficulties, social distancing, being under lockdown, quarantine measures, and working remotely. With people heavily relying on social media platforms to share information more than ever, it is important to analyze their conversations to understand people's sentiments and feelings during this time of crisis to find possible ways to cope with the pandemic. This paper presents a sentiment analysis study to analyze sentiments from Arabic tweets related to COVID-19 using multiple models. After data acquisition, text preprocessing steps are performed and Term Frequency Inverse Document Frequency (TF-IDF) is used to generate feature vectors. Experiments are then done comparing multiple classifiers: Naïve Bayes, Support Vector Machine, Logic Regression, Random Forest, and K-Nearest Neighbor. A comparison of the models' performance was carried out using multiple evaluation metrics including Precision, Accuracy, Recall and F1 Score. The best performing model achieved an accuracy of around 84%.

Keywords Sentiment analysis · Arabic · Tweets · Twitter · COVID-19 · Coronavirus

D. Ahmed · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_50

1 Introduction

Nowadays, social media networks have become a useful and remarkable source of news, information and people's reactions and opinions regarding events. With people freely expressing their feelings and opinions on social media platforms, this gives interested organizations a great chance to benefit from this data [1]. Twitter is the most popular social media platform utilized to access people's feelings since they have over 300 million active users per month [2]. This facilitates the process of investigating people's opinions and feelings since there is a huge amount of available data. This analysis process is called Sentiment Analysis (SA), where a sentiment can be defined as a thought or attitude caused by a feeling.

Most of the available work on SA is focused on the English language and a much less percentage is done on the Arabic language. Over the past few years, Arabic Sentiment Analysis (ASA) has seen a significant increase in the researchers' interest because of the exponential increase in the number of Arabic content on social media (specifically Arabic tweets) across various domains such as economics, reviews, online marketing, communities and many more [3]. Arabic is a morphologically rich and derivational language with unique characteristics, several dialect variations, multiple forms for words, and significant use of diacritical symbols compared to other languages. This makes the process of SA more complex when dealing with Arabic content [4].

There have been many studies exploiting tweets to perform SA on crisis situations in many languages such as [5–7]. Moreover, analysis of tweets has been done before in cases of health-related crises. Many research efforts have been done exploiting tweets to identify disease outbreaks such as the popular case of H1N1 virus [5] and influenza [6]. In addition, it helped with recognizing people's feelings about other outbreaks such as Ebola [3] and Zika virus [7]. So, this can be extended to understand people's sentiments about the recent COVID-19 pandemic. With the increase of COVID-19 cases in the Arab world and the presence of enormous amounts of related conversations on social media platforms such as Twitter, the need to understand people's feelings to provide the best help and support is growing stronger. Since the start of the pandemic and the first COVID-19 case followed by the huge infection rates all over the world, social media engagement kept increasing. When new restrictions were applied such as social distancing, quarantine, face masks, shifting to remote working, closing public places and people losing their jobs, people tweeted all kinds of different information about their situation and tried to always check updates from all trusted accounts for healthcare officials and governmental entities. At that time, it would have been useful to have analysis results for how the people were feeling at each of those stages where new updates arose so that the corresponding organizations can help them. For instance, government entities can check if the public is satisfied with new regulations applied and then try to accommodate any dissatisfaction if the general sentiment was negative. Also, we can analyze how closures were affecting students' studies and communicate these findings with educational institutions to create new plans to

make sure the students' future careers won't be negatively impacted. Moreover, charities or aid organizations can provide help to people expressing negative sentiments over losing jobs or in need of financial assistance because of COVID-19 repercussions. Sick or old individuals can also be mapped with volunteers and medical assistance in case of emergency tweets expressing negative emotions.

Since COVID-19 is a relatively new pandemic, the amount of SA studies done addressing the topic are not in abundance, especially the studies using Arabic social media data. It is important to utilize the enormous amount of social media data and closing the research gap by contributing to the ongoing ASA research. Therefore, this paper aims to perform SA on Arabic tweets discussing COVID-19 to understand people's sentiments to help assist corresponding health organizations and scientists. Several machine models will be experimented with and their performance will be compared to each other to determine the best performing model in the SA process.

This paper is divided into the following sections. Section 2 presents some of the related work on SA, ASA, and COVID-19. Next, Sect. 3 explains the research methodology with experimental details. Section 4 then demonstrates the results of the experiments and discusses them. Section 5 concludes.

2 Related Work

2.1 SA in Crisis Situations

Data from Twitter has been used for analyzing people's sentiments in crisis situations such as pandemics and disease outbreaks [8–10]. The authors in [11] analyzed the temporal progress of diseases connected to influenza in New York City in the US through tweets posted on the topic. In [12], the authors performed SA on a Spanish corpus containing tweets on infectious illnesses in America such as Zika and Dengue viruses. The corpus had over 10,000 positive tweets, around 7000 neutral tweets and over 10,000 negative tweets. They also included linguistic elements in the feature engineering and added word embedding to handle the aspect SA process. Another study on influenza [13] performed a SA on over 50,000 Arabic tweets talking about influenza spread across several countries. Different ML algorithms were used including Decision trees (DT), Support Vector Machine (SVM), K-nearest neighbor (k-NN), and Naïve Bayes (NB). The k-NN performed the best over tweets from the Gulf region, the NB algorithm performed the best over tweets from Syria, and SVM performed the worst for tweets from most countries.

Moreover, one of the first Arabic corpora tackling crisis data is presented in [14], where it contains tweets about the starvation crisis in Yemen, cholera epidemic outbreak, and the issue of Syrian refugees. The researchers collected the Arabic tweets regarding these crisis topics, annotated the data, preprocessed the text, extracted needed features, built classification models and then evaluated them.

A comparison between the ML models (Naïve Bayes NB, Support Vector Machine SVM and Random Forest Classifier RF) was done to test for the prediction of the Arabic tweets' labels.

In [10], both the learning and lexicon-based approaches were tested on a dataset of Arabic tweets on the Syrian refugee crisis and the civil war there as well. The experiment results showed that the learning-based model with logistic model trees had the best performance after a comparison amongst several ML algorithms (logistic regression, SVM, decision trees, K-Nearest Neighbor). Furthermore, the authors in [15] did a SA study to utilize tweets to see the geographical distribution of health issues related to consumed food in different cities in the US, showing that regions, where people have low income, exhibited more health issues because of unhealthy food consumption.

2.2 SA During COVID-19

Authors in [16] obtained over 500,000 tweets using hashtags related to COVID-19, corona, and coronavirus. Their experiments resulted in over around 50% of neutral tweets, over 30% positive sentiments tweets, and a little above 10% for negative tweets. Another study [17] using the same hashtags but focused in Nepal on a much smaller amount of tweets (a little above 600) achieved results of 30% neutral tweets, 14% negative tweets and 60% positive tweets. Furthermore, a SA focused in India was done in [18] on around 25,000 tweets resulted in a majority of positive sentiments.

The study in [19] performed a similar experiment using the same hashtags COVID-19 and coronavirus. The tweets with COVID-19 tags had a majority of positive sentiments (over 50%), followed by neutral sentiments of a little over 30%, and only a bit over 10% represented negative sentiments. The other tag "coronavirus" had the highest percentage of neutral sentiments (over 40%) closely followed by positive sentiments (over 35%) and a bit over 15% for negative tweets. All the studies done showed similar results where the people's sentiments and feelings regarding the COVID-19 pandemic were mostly positive or neutral.

A recent study was done in [20] to examine the sentiments of people in Saudi Arabia to the preventive measures applied during COVID-19 by analyzing Arabic tweets. They collected over 50,000 tweets matching different points of time corresponding to applied governmental measures such as lockdown of mosques, cancellation of group prayers in mosques, closing of schools, universities, malls, and restaurants. The experiment resulted in a majority of positive sentiments showing that the population was happy with the health precautions implemented by the government. Another study [21] develop an emotion detection approach to classify COVID-19 Arabic tweets into eight emotions.

The research body on ASA addressing COVID-19 is increasing but still lacking in comparison to other SA domains. Most of the studies done are targeting a specific country audience like Nepal [17] or Saudi Arabia [21] handling a limited dataset restricted by geographical restrictions. This study uses a much larger dataset that is

more diverse over several Arab countries instead of a limited sample. Also, many of the studies just use one or two machine learning models for the SA with conflicting results across papers, but this study will compare the performance of five different models to determine which one performs best when it comes to ASA on COVID-19 tweets. Moreover, most studies perform a partial preprocessing procedure since Arabic data preprocessing is more complex compared to English, but this study will perform all detailed preprocessing techniques to ensure the quality of the dataset used.

3 Methodology

This section describes the experimental methodology adopted for this sentiment analysis study and demonstrates the steps followed to identify the sentiments in Arabic tweets regarding COVID-19. First, the data is acquired through a training and a test dataset. The collected data is then preprocessed to have the data in an appropriate form, followed by feature extraction to construct the classification model. Multiple classification models are used to classify the tweets into sentiment classes of positive or negative, and then the results are compared to identify the best performing classification model.

3.1 Datasets

A labelled dataset is needed to develop the ML model that would be able to process the tweets and predict their classification. The dataset used in this study is "Arabic Sentiment Twitter Corpus (ASTC)" [22]. It contains 58,000 Arabic tweets that were gathered in April 2019. The tweets are divided into 47,000 tweets for training and 11,000 tweets for testing, and they are all annotated with positive and negative sentiment labels. The labels were done using positive and negative emojis lexicon. The division of the dataset is balanced since the percentage of positive and negative tweets is the same, with each amounting to 50% [22]. To analyze people's sentiments and feelings about COVID-19, the "Arabic Twitter COVID-19 Collection" dataset [23] is used. This has almost 4 million Arabic tweets related to COVID-19. The authors collected the data through Twitter's streaming API for the time duration from January 2020 till the beginning of May 2020. The dataset is publicly available on the authors' GitHub page [24].

3.2 Pre-processing Stage

In In tweets text mining, it is important to prepare the data in an optimal form that would be appropriate for analysis. This data preprocessing includes multiple stages in

order to clean the data and reduce noise. For this study, the following text preprocessing steps were applied. These are a mixed collection of steps obtained from work done by other researchers in [25, 26] to cover most of the necessary operations.

- Cleaning: Any unnecessary text will be removed from the tweets. This includes numbers, special characters or punctuation marks such as ($, |, &, %, -, _, “”; ‘, etc.…), white spaces, hashtags (#something), mentions (@usernames), emojis, and URL links. Also, any letters written in a language other than Arabic were removed. Duplicate tweets, retweets and short tweets with only one to three words were filtered out as well.
- Removal of the diacritical symbols (tashkeel تشكيل): These symbols are marks below or above the individual letters that give different pronunciations and meanings based on the grammar and syntax of the sentence. They are removed to get a unified shape for the words.
- Tokenization of each word in the tweets to generate a group of tokens as an array consisting of strings
- Removal of stop words: Words that don’t have meaning or don’t change the meaning of the text such as articles, conjunctions and prepositions are considered stop wards and are removed. In English, these could be “at, the, an, in, on, a, etc.…. Some examples of stop words in Arabic are: (في, على, الى, من, الـتي, الكل, من, ما, ما, قد, لم)
- Removing repeated letters: There are frequent cases in social media posts (tweets in this case) where words there are letters that are repeated to emphasize some meaning or feeling. For example, “الحقووووووووونى”, where the letter “و” is repeated, so it would be removed to become “الحقونى” which is in Egyptian dialect translating to a plea of help “Save me”.
- Normalization: This is done to transform the different forms of a single letter to a uniform letter across all tweets.
  - For instance, there are different forms of the letter “Alef” so (أ, إ, آ) would be replaced by “ا”. In addition, replacing “ة” with “ه”, replacing “ؤ” with “و”, replacing “ئ” with “ي”, and replacing “ڤ” (v sound) with “ف” (f sound) especially because the word “virus” when written in Arabic can be represented as either “فيروس” or “ڤيروس”.
  - In Arabic, the word “virus” could be pronounced as either “Firus”, “Virus”, “Fairus”, or “Vairus” depending on different accents and dialects. So, words like “فايروس” would be replaced “فيروس”.
  - Removing elongations of letters in the words. For instance, replacing “فـــــــيروس” with “فيروس”.
- Stemming: Any compound words would be replaced by the stem/root. The stem/root would be extracted from the word tokens created in the tokenization step. For example, in English, the words “player”, “playing”, and “played” are have the same root which is the word “play”. This is important to minimize repetition. The IRSI Arabic stemmer from the NLTK python package is used for stemming the Arabic data [27] (Table 1).

Table 1 Example processing steps on a COVID-19 tweet

Original tweet	@kuwait_news إصاااااااااابة 7 مُمرضيـــــن و أطِبـَـاء في مُسْتَشْفَي ال Amiri بِڤايرووووس كورونا و إشتباه في 5 حالاااااااااات مُخَالـــــطة لَـــــهُم https://alqabas.com/article/5797996- #الكويت #كورونا Translation in English: Seven nurses and doctors in the Amiri hospital have been diagnosed with corona virus and 5 suspected cases
After cleaning	إصَاااااااااابة مُمرضيـــــن و أطِبـَـاء في مُسْتَشْفَي بِڤايرووووس كورونا و إشتباه في حالاااااااااات مُخَالـــــطة لَـــــهُم Translation in English: Several nurses and doctors in hospital have been diagnosed with corona virus and multiple suspected cases
After removing diacritical symbols	إصاااااااااابة ممرضيـــــن و أطبـــاء في مستشفي بِڤايرووووس كورونا و إشتباه في حالاااااااااات مخالـــــطة لَـــــهم Translation in English: Several nurses and doctors in hospital have been diagnosed with corona virus and multiple suspected cases
After removing repeated letters	إصابة ممرضيـــــن و أطبـــاء في مستشفي بِڤايروس كورونا و إشتباه في حالات مخالـــــطة لَـــــهم Translation in English: Several nurses and doctors in hospital have been diagnosed with corona virus and multiple suspected cases
After removal of stop words	إصابة ممرضيـــــن أطبـــاء مستشفي ڤايروس كورونا إشتباه حالات مخالـــــطة Translation in English: Several nurses doctors hospital have been diagnosed corona virus multiple suspected cases
After normalization	اصابه ممرضين اطباء مستشفي فيروس كورونا اشتباه حالات مخالطه Translation in English: Several nurses doctors hospital diagnosed corona virus multiple suspected cases

3.3 Feature Extraction

After the data preprocessing, we have a set of bag of words representing the original feature vector, which will be utilized to generate the feature vector. The most commonly used technique to do this is using Term Frequency-Inverse Document Frequency (TF-IDF), which assigns a degree of importance to each word in a document (a tweet in this case) [28]. Each word would be assigned some weight representing the importance which will increase when the number of times the word appears in the document/tweet increases while being offset by the overall frequency of the word in the dataset [12].

4 Results and Discussion

Multiple classification models are applied and tested over the two datasets [22, 23]. A comparison was performed for the performance of the various models. Each of the datasets is randomly distributed into testing and training subsets. Table 2 shows the positive and negative sentiments distribution for the datasets. As mentioned in Sect. 3.1, the first dataset was balanced with an equal representation of positive and negative sentiments. The second dataset has around 56% positive sentiments and 44% negative sentiments.

The majority of the sentiments classified were found to be positive. This might not be entirely an accurate representation of people's sentiments as there could be underlying elements affecting their posts. Some individuals might be reluctant to freely say their negative feelings and opinions on a public social media platform since it might look bad or be considered as offensive and insensitive to the continuous efforts done by governments and health organizations to help face the COVID-19 pandemic.

The data is classified into two sentiment classes either positive or negative. Table 3 shows the results for the performance evaluation measures: accuracy, precision, recall and F1 score across five different machine learning models.

As shown in Table 3, the four models SVM, NB, RF and LR all performed well while achieving close results. SVM got the highest accuracy, which of 84.14%, closely followed by NB with accuracy of 82.67%, followed by RF with accuracy of 81.74% and then LR with accuracy of 80.46%. K-NN has the lowest accuracy of 63.23%. In addition, Support Vector Machine has the highest values for precision and recall, while k-NN has the lowest values in all four performance metrics. The low results for the k-NN model can be attributed to the fact that it assigns equal weights to each individual attribute in the feature vector, which can result in confusion and ambiguities if there is a significant number of irrelevant attributes present. In addition, the k-NN model is built on using similarity measures by

Table 2 Tweets distribution over sentiment classes

Sentiment class	Dataset 1 (%)	Dataset 2 (%)
Positive	50	56.41
Negative	50	43.59

Table 3 ML models performance validation

ML classification model	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)
LR	80.46	79.71	80.45	71.36
SVM	84.14	85.80	86.40	86.06
NB	82.67	82.62	84.99	89.77
RF	81.74	81.98	80.99	82.82
k-NN	63.23	41.65	36.74	38.92

measuring the distance between the tweets by applying several distance measures like the Euclidean distance. This looks only at the distance and ignores the semantic relations between the tweets. This might have led to the misclassification of some tweets. This justifies the low-performance measures compared to the other classification models.

5 Conclusion

The study in this paper tackles the task of comparing models that can predict people's sentiments from Arabic tweets throughout the COVID-19 pandemic outbreak. To achieve this, the tweets are cleaned and preprocessed using several NLP techniques to obtain the tweets that represent the dataset the best in the appropriate optimal format. TF-IDF is used to generate the feature vector, and then multiple classifiers were used to check the performance of the different ML models. Five models were compared which were LR, SVM, RF, NB and K-NN. To measure the performance, four evaluation metrics were reported: Precision, Accuracy, Recall and F1 Score. The experimental results in this study show that SVM, NB, RF and LR performed well with close results, while SVM had the best performance with an accuracy of 84.14% and also got the best results in terms of the true positive rate with a recall of 86.40%. The worst result was obtained by the k-NN model with an accuracy of 63.23% and also the worst performance in terms of the true positive rate with a recall of 36.74%.

References

1. Al-Twairesh N, Al-Khalifa H, Alsalman A, Al-Ohali Y (2018) Sentiment analysis of Arabic tweets: feature engineering and a hybrid approach
2. Hamshere T (2013) Getting started with Twitter flight. Packt Publishing Ltd
3. Morin C, Bost I, Mercier A, Dozon J-P, Atlani-Duault L (2018) Information circulation in times of Ebola: Twitter and the sexual transmission of Ebola by survivors. PLoS Curr 10. https://doi.org/10.1371/currents.outbreaks.4e35a9446b89c1b46f8308099840d48f
4. Alwakid G, Osman T, Hughes-Roberts T (2017) Challenges in sentiment analysis for Arabic social networks. Procedia Comput Sci 117:89–100. https://doi.org/10.1016/j.procs.2017.10.097
5. De Quincey E, Kostkova P (2010) Early warning and outbreak detection using social networking websites: the potential of Twitter. In: Lecture notes of the institute for computer sciences, social-informatics and telecommunications engineering, LNICST, vol 27, pp 21–24. https://doi.org/10.1007/978-3-642-11745-9_4
6. Culotta A (2010) Towards detecting influenza epidemics by analyzing Twitter messages. In: SOMA 2010—Proceedings of the 1st workshop on social media analytics, pp 115–122. https://doi.org/10.1145/1964858.1964874
7. Glowacki EM, Lazard AJ, Wilcox GB, Mackert M, Bernhardt JM (2016) Identifying the public's concerns and the centers for disease control and prevention's reactions during a health crisis: an analysis of a Zika live Twitter chat. Am J Infect Control 44(12):1709–1711. https://doi.org/10.1016/j.ajic.2016.05.025

8. Fung I et al (2019) Pedagogical demonstration of Twitter data analysis: a case study of world AIDS day, 2014. Data 4(2):84. https://doi.org/10.3390/data4020084
9. Kim EH-J, Jeong YK, Kim Y, Kang KY, Song M (2016) Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news. J Inf Sci 42(6):763–781. https://doi.org/10.1177/0165551515608733
10. Samuel J, Rahman MM, Ali GGMN, Samuel Y, Pelaez A (2020) Feeling like it is time to reopen now? COVID-19 new normal scenarios based on reopening sentiment analytics. SSRN Electron J. https://doi.org/10.2139/ssrn.3604802
11. Nagar R et al (2014) A case study of the New York City 2012–2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. J Med Internet Res 16(10):e236.https://doi.org/10.2196/jmir.3416
12. García-Díaz JA, Cánovas-García M, Valencia-García R (2020) Ontology-driven aspect-based sentiment analysis classification: an infodemiological case study regarding infectious diseases in Latin America. Futur Gener Comput Syst 112:641–657. https://doi.org/10.1016/j.future.2020.06.019
13. Baker QB, Shatnawi F, Rawashdeh S, Al-Smadi M, Jararweh Y (2020) Detecting epidemic diseases using sentiment analysis of arabic tweets. J Univers Comput Sci 26(1):50–70
14. Adel G, Wang Y (2019) Arabic twitter corpus for crisis response messages classification. In: ACM international conference proceeding series, pp 498–503. https://doi.org/10.1145/3377713.3377799
15. Widener MJ, Li W (2014) Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. Appl Geogr 54:189–197. https://doi.org/10.1016/j.apgeog.2014.07.017
16. Manguri KH, Ramadhan RN, Mohammed Amin PR (2020) Twitter sentiment analysis on worldwide COVID-19 outbreaks. Kurdistan J Appl Res:54–65. https://doi.org/10.24017/covid.8
17. Pokharel BP (2020) Twitter sentiment analysis during Covid-19 outbreak in Nepal. SSRN Electron J. https://doi.org/10.2139/ssrn.3624719
18. Barkur G, Vibha, Kamath GB (2020) Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: evidence from India. Asian J Psychiatr 51:102089. https://doi.org/10.1016/j.ajp.2020.102089
19. Bhat M, Qadri M, Beg N, Kundroo M, Ahanger N, Agarwal B (2020) Sentiment analysis of social media response on the Covid19 outbreak. Brain, Behav Immunity 87:136–137. https://doi.org/10.1016/j.bbi.2020.05.006
20. Alhajji M, Al Khalifah A, Aljubran M, Alkhalifah M (2020) Sentiment analysis of tweets in Saudi Arabia regarding governmental preventive measures to contain COVID-19. https://doi.org/10.20944/preprints202004.0031.v1
21. Alhazmi H, Alharbi M (2020) Emotion analysis of Arabic tweets during COVID-19 pandemic in Saudi Arabia
22. Saad M (2019) Arabic sentiment twitter corpus: positive and negative tweets collected from twitter. Available: https://www.kaggle.com/mksaad/arabic-sentiment-twitter-corpus
23. Alqurashi S, Alhindi A, Alanazi E (2020) Large Arabic twitter dataset on COVID-19. arXiv, Apr 2020
24. Alqurashi S, Alhindi A, Alanazi E (2020) COVID-19 Arabic tweets dataset. Available: https://github.com/SarahAlqurashi/COVID-19-Arabic-Tweets-Dataset
25. Al-kabi MN, Gigieh AH, Alsmadi IM, Wahsheh HA (2014) Opinion mining and analysis for Arabic language. Int J Adv Comput Sci Appl 5(5):181–195
26. Ghallab A, Mohsen A, Ali Y (2020) Arabic sentiment analysis: a systematic literature review. Appl Comput Intell Soft Comput 2020. https://doi.org/10.1155/2020/7403128
27. Abuelenin S, Elmougy S, Naguib E (2018) Twitter sentiment analysis for arabic tweets. Adv Intell Syst Comput 639:467–476. https://doi.org/10.1007/978-3-319-64861-3_44
28. Ye X, Li S, Yang X, Qin C (2016) Use of social media for the detection and analysis of infectious diseases in China. ISPRS Int J Geo-Information 5(9):156. https://doi.org/10.3390/ijgi5090156

Psychometric Characteristics of the Iranian Smartphone Addiction Inventory Short Form (SPAI-SF)

Ibrahim Arpaci, Saeed Barzegari, Hamed Mahmoodi, Elham Afrooz, and Avishan Zanganeh Ranjbar

Abstract This study aimed to develop and validate a short form of the Smartphone Addiction Inventory (SPAI) and determine the cut-off point for smartphone addiction based on a psychiatric interview. A sample of 281 Iranian undergraduate students from Mazandaran University of Medical Sciences completed the 26-item SPAI and psychologist interview. Stepwise multiple linear regression identified 10 items for the short-form of the SPAI, consisting of compulsive behavior (four items), functional impairment (three items), withdrawal (two items), and tolerance (one item) with an explained variance of 96%. Exploratory factor analysis revealed a two-factor structure that labeled Compulsive behavior and Functional impairment (47.99 variance) and Withdrawal and Tolerance (11.57 variance). Confirmatory factor analysis (CFA) verified the structure by excellent goodness of fit indices (p-value = 0.061, x^2/df = 1.43, CFI = 0.99, GFI = 0.97, and RMSEA = 0.039). The inter-item and item-total correlations of the SPAI-SF were positively significant and confirmed the convergent validity. Internal consistency verified by Cronbach's alpha of 0.875 (CI: 0.852 − 0.896). The area under the curve (AUC) was 0.876 (95% CI = 0.833 − 0.919, $p < 0.001$), indicating a good diagnostic efficiency. The best cut-off point to discriminate diagnostic negatives was 23, with a sensitivity of 79.19%, a specificity of 83.33%, and an accuracy of 80.43%. The results indicated acceptable reliability, validity, and AUC for the SPAI-SF for smartphone addiction screening and studies.

Keywords Smartphone addiction · Cut-off point · SPAI-SF

I. Arpaci
Tokat Gaziosmanpasa University, 60250 Tokat, Turkey

S. Barzegari (✉) · E. Afrooz · A. Z. Ranjbar
Mazandaran University of Medical Sciences, Sari, Iran

H. Mahmoodi
Kashan University Medical Sciences, Kashan, Iran

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_51

1 Introduction

Smartphones are digital tools with the capabilities of both computers and mobiles, for this reason, its use has become global [1]. However, spending a long time with a smartphone have complications such as smartphone addiction [2]. Smartphone addiction is considered as "a behavioral addiction that involves human–machine interaction and is non-chemical in nature" [3]. Addiction can be defined as a mental disorder that shows symptoms of withdrawal, dependence, tolerance along with problems in time management [4]. Addiction covers both substance dependency as well as behavioral addictions (e.g., social media, Internet, smartphone, and online gaming) [5]. However, the use of the term "smartphone addiction" has not been used by some experts in the literature. Therefore, in this study "smartphone addiction" means "excessive internet use" or "problematic internet use."

Chen et al. identified the drivers of the smartphone addiction. Their results indicated that mood regulation, conformity, pastime, and perceived enjoyment were positively associated with smartphone addiction [6]. Bock et al. proposed a measure to explain factors related to mobile phone addiction. Their results indicated that mobile phone addiction is related to depression and impulsivity [7]. In another study, Kim et al. revealed that impulsivity is significantly related to smartphone addiction [8]. Lin et al. indicated that "compulsive behavior," "functional impairment," "withdrawal," and "tolerance" were significantly related to smartphone addiction [9].

Developing a brief, valid, and reliable measurement with a good discrimination between addicted and non-addicted person is an important research topic. Accordingly, the study aimed to develop the short form of the SPAI in Iran. Thus, the study investigated validity and reliability of the SPAI-SF. Further, the study established a cut-off value for the short-form SPAI (SPAI-SF) using diagnostic criteria resulted from psychiatric interviews.

2 Method

2.1 Sample and Procedure

Sample of the study was 260 Iranian undergraduate students in Nursing, Health Information Technology, and Laboratory Sciences disciplines (122 males, 148 female) from the Mazandaran University of Medical Sciences (MAZUMS). They completed the SPAI during May to October in 2019. The mean age of students was 20.12 ± 2.01 years. Students that were smartphone users and completed informed consent were included in the study. Furthermore, each participant individually interviewed by a psychiatrist. The psychiatrists, as the gold standard, evaluate participant's smartphone usage.

2.2 Measures

The SPAI is a 26 item inventory and scored on a four-point Likert format from "strongly disagree" to "strongly agree." Total score of the SPAI ranges from 26 to 104. Four components identified as: "Compulsive behavior, Functional impairment, Withdrawal, and Tolerance." SPAI had an adequate internal reliability consistency ($\alpha = 0.94$) [9].

3 Results

The most common approaches to selecting items and develop short form of scales are the correlation, stepwise multiple linear regression, and factor analysis. To form candidate items of the SPAI-SF, stepwise multiple linear regression was conducted. Items with the highest explained variance of the sum score of the SPAI-26 were selected as candidates for the short form. The construct validity was tested using inter-item, item-scale (SPAI-SF), and item-scale (SPAI-26) correlations, Exploratory Factor Analysis (EFA), and CFA, which was performed via SPSS and AMOS (ver.20). Promax rotation was used in the EFA because of the high correlation between the items. The factors that have eigenvalues greater than one were retained for the CFA. KMO and Bartlett's Test of Sphericity indices were checked for the adequacy of sample size, and the explained variance by items was calculated.

In the next step, model fit indices were evaluated via examining the "comparative fit index (CFI), incremental fit index (IFI), and root-mean-square error of approximation (RMSEA), Chi-Square/degree of freedom (x^2/df), and Goodness-of-fit index (GFI)." We also calculated the Cronbach's α to evaluate the internal reliability consistency.

Receiver operating characteristics (ROC) was used to calculate the area under the ROC curve (AUC) and examine the diagnostic validity of the scale. The specificity, sensitivity, PPV, NPV, Youden Index, Correctly Classified, LR+, and LR- scores were used to determine diagnostic negative and positive groups. The best cut-off point of the SPAI-SF was selected by a high Youden Index and maximal diagnostic accuracy. The demographic characteristics and smartphone usage were compared between these two groups by the chi-square test or t-test.

Multiple-linear-regression was used to identify the items which had the highest explained variance for the SPAI-26. The dependent variable was the total score of the SPAI-26 and the independent variables were the scores of each item. The stepwise multiple linear regression employed to figure out the best predictors of the total score. Items with the highest contribution to the coefficient of variation (adjusted R^2) were chosen for the SPAI-SF. Results showed that 10 items explained 95.7% of the total variance or adjusted R2 and had good representatives for the SPAI-26. The SPAI-SF included three items (Q17, Q23, Q24) from "Functional

Impairment," four items (Q7, Q10, Q11, and Q22) from "Compulsive Behavior," two items (Q16, Q25) from "Withdrawal," and one item (Q3) from "Tolerance."

The item-total correlations of the SPAI-SF were statistically significant ($p < 0.001$). The highest correlations were for items 7(0.77), 11(0.77), 23(0.74), 10 (0.74), and 22(0.73), respectively. The alpha for the SPAI-SF was 0.875 (CI: 0.852-0.896). First factor 0.884 (0.863-0.904), second factor 0.633 (0.552-0.701). Table 1 shows that the correlations of the total score of SPAI-26 and items of the SPAI-SF were statistically significant ($r = 0.98$, $p < 0.001$).

All the items have a significant correlation at the 0.01 level (2-tailed). The inter-item correlation coefficient (0.15–0.67) was computed for checking the multicollinearity (See Table 2). The EFA with Promax rotation extracted two factors

Table 1 Explained variance by each item in stepwise multiple linear regression, and the inter-item and item-total correlations

Step	Item added	R	R^2	Adjusted R^{2*}	SPAI *r*	SPAI-SF *r*
1	Q23: "I make it a habit to use smartphone and the sleep quality and total sleep time decreased"	0.77	0.59	0.59	0.77	0.74
2	Q11: "I fail to control the impulse to use smartphone"	0.87	0.76	0.76	0.74	0.77
3	Q25: "I cannot have meal without smartphone use"	0.91	0.83	0.83	0.59	0.63
4	Q7: "Although using smartphone has brought negative effects on my interpersonal relationships, the amount of time spent on Internet remains unreduced"	0.95	0.87	0.87	0.76	0.77
5	Q17: "My interaction with family members is decreased on account of smartphone use"	0.95	0.90	0.90	0.70	0.71
6	Q3: "I find that I have been hooking on smartphone longer and longer"	0.96	0.92	0.92	0.39	0.42
7	Q10: "I feel distressed or down once I cease using smartphone for a certain period of time"	0.97	0.93	0.93	0.72	0.74
8	Q24: "I need to spend an increasing amount of time on smartphone to achieve same satisfaction as before"	0.97	0.94	0.94	0.72	0.71
9	Q22: "I try to spend less time on smartphone, but the efforts were in vain"	0.98	0.95	0.95	0.72	0.73
10	Q16: "I feel missing something after stopping smartphone for a certain period of time"	0.98	0.96	0.96	0.56	0.60

*Percent of explained variance of the total scale score

Table 2 Factor loadings and inter-item correlations of the SPAI-SF

Factor (Explained variance)	Item	Load	Communality	Q7	Q 10	Q 11	Q 17	Q 22	Q 23	Q 24	Q 3	Q 16	Q 25
Compulsive behavior and functional impairment (47.99)	7	0.76	0.63	1	0.56	0.61	0.47	0.59	0.55	0.51	0.23	0.35	0.36
	10	0.72	0.58		1	0.67	0.42	0.55	0.48	0.47	0.17	0.32	0.43
	11	0.67	0.62			1	0.42	0.56	0.50	0.45	0.22	0.42	0.43
	17	0.67	0.51				1	0.42	0.58	0.55	0.20	0.30	0.34
	22	0.74	0.57					1	0.54	0.47	0.25	0.40	0.37
	23	0.91	0.69						1	0.66	0.15	0.30	0.28
	24	0.84	0.61							1	0.18	0.28	0.34
Withdrawal and tolerance (11.57)	3	0.77	0.49								1	0.32	0.26
	16	0.79	0.66									1	0.52
	25	0.69	0.60										1

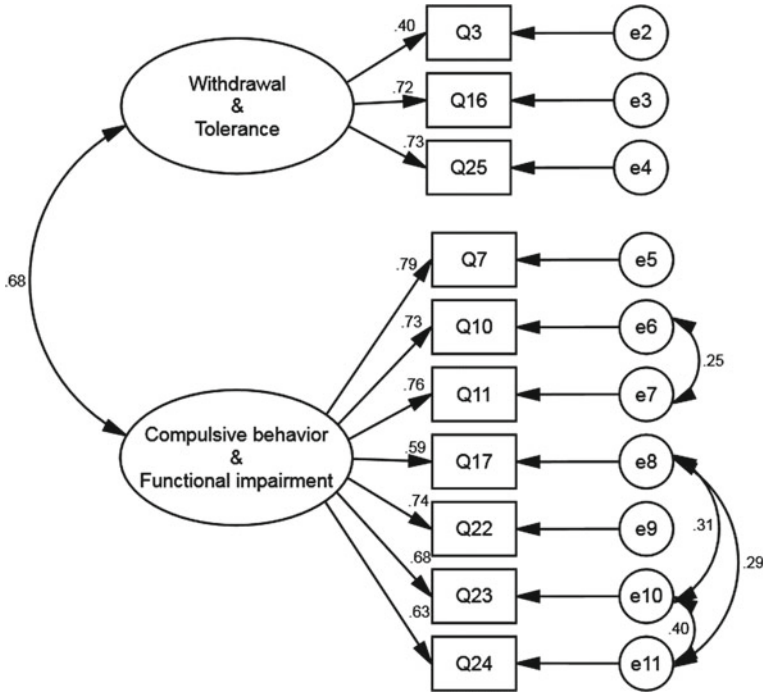


Fig. 1 The fitted model for the SPAI-SF

Labeled: Cb-Fi and WT (59.57% of the total variance). Communality of items was between 0.49 and 0.69. The KMO value was 0.894, and Bartlett's test of sphericity was also significant (df = 45, x^2 = 1176.3, $p < 0.001$).

Outliers were examined by the Mahalanobis distance, no outliers were found. The skewness and kurtosis of the items were tested to verify the normality of distribution. Figure 1 indicated the SPAI-SF and the factor loading of each item, based on a two-factor structure. The factor loadings were ranged from 0.40 to 0.79. The indices were also acceptable (p-value = 0.061, df = 30, x^2 = 42.77, x^2/df = 1.43, CFI = 0.99, GFI = 0.97, AGFI = 0.95, IF = 0.99, TLI = 0.98, and RMSEA = 0.039).

Among the 281 participants, 84 were classified as "diagnostic positive" group, while 197 were classified as "diagnostic negative" group. Figure 2 shows the ROC curve for the SPAI-SF designed by sensitivity and 1-specificity.

The AUC was 0.876 (95% CI = 0.833–0.919, SE = 0.022, $p = \leq 0.001$), indicating a good diagnostic efficiency. Table 3 indicates that a cut-off point of 23 was the best for discriminating addicted cases from diagnostic negatives (correctly classified: 80.43%, Youden Index: 0.63, specificity: 79.19%, and sensitivity: 83.33%).

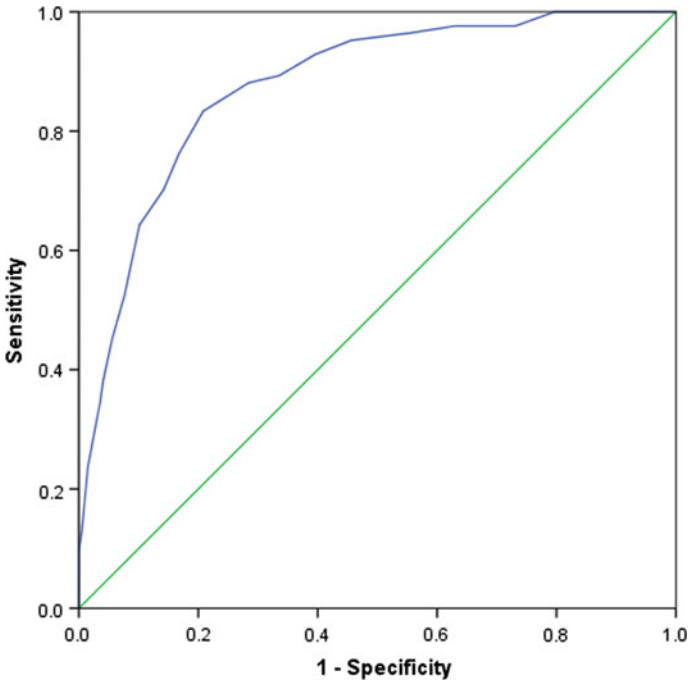


Fig. 2 ROC curve for the SPAI-SF

Table 3 The cut-off point, sensitivity, specificity, and classification accuracy of the SPAI-SF

Cut-off point	Sensitivity %	Specificity %	Correctly classified	Youden index	LR+	LR−
0.09	2.08	0.496	66.55	54.31	95.24	≥ 19
0.12	2.35	0.533	70.11	60.41	92.86	≥ 20
0.16	2.67	0.558	73.31	66.50	89.29	≥ 21
0.17	3.10	0.597	76.51	71.57	88.10	≥ 22
0.21	**4**	**0.625**	**80.43**	**79.19**	**83.33**	**≥ 23**
0.29	4.55	0.594	81.14	83.25	76.19	≥ 24
0.35	4.94	0.560	81.14	85.79	70.24	≥ 25
0.40	6.33	0.541	82.21	89.85	64.29	≥ 26
0.52	6.88	0.448	80.43	92.39	52.38	≥ 27
0.58	8.10	0.397	79.72	94.42	45.24	≥ 28
0.65	9.38	0.340	78.65	95.94	38.10	≥ 29
0.68	9.72	0.310	77.94	96.45	34.52	≥ 30

4 Discussion and Conclusion

This study aimed to develop and validate a short-form of the SPAI in Iranian undergraduate students and determine the cut-off point for screening smartphone addiction. The results of this study indicated adequate validity and reliability for the SPAI-SF. Previous studies also indicated that the SPAI is a valid and reliable scale to measure individuals' smartphone addiction level [10, 11]. The present study validated a 10-item self-report instrument with a four-point Likert-scale to measure the levels of smartphone addiction. The results suggested 10 items with two-factors: "withdrawal and tolerance" and "compulsive behavior and functional impairment". The results showed that the SPAI-SF demonstrates good discriminant and convergent validity along with a strong internal consistency. Further, the results showed that cut-off point of 23 was the best for discriminating addicted cases from diagnostic negatives.

In conclusion, the findings have provided a strong evidence that the SPAI-SF has promising validity and reliability properties. Nonetheless, this study has certain limitations that need to be addressed. The samples were non-diagnosed subjects in terms of addiction. The measurement should therefore be tested with clinically diagnosed subjects.

Funding This work was supported by the Mazandaran University of Medical Sciences with code IR.MAZUMS.REC.1398.618.

References

1. Arpaci I, Cetin Yardimci Y, Turetken O (2015) A cross-cultural analysis of smartphone adoption by Canadian and Turkish organizations. J Glob Inf Technol Manag 18(3):214–238. https://doi.org/10.1080/1097198X.2015.1080052
2. Chiang J-T, Chang F-C, Lee K-W, Hsu S-Y (2019) Transitions in smartphone addiction proneness among children: the effect of gender and use patterns. PloS One 14(5). https://doi.org/10.1371/journal.pone.0217235
3. Lin YH, Pan YC, Lin SH, Chen SH (2017) Development of short-form and screening cutoff point of the smartphone addiction inventory (SPAI-SF). Int J Meth Psychia Res 26(2):e1525. https://doi.org/10.1002/mpr.1525
4. O'Brien C (2011) Addiction and dependence in DSM-V. Addiction 106(5):866–867. https://doi.org/10.1111/j.1360-0443.2010.03144.x
5. Young KS (1998) Internet addiction: the emergence of a new clinical disorder. Cyberpsychol Behav 1(3):237–244. https://doi.org/10.1089/cpb.1998.1.237
6. Chen C, Zhang KZ, Gong X, Zhao SJ, Lee MK, Liang L (2017) Examining the effects of motives and gender differences on smartphone addiction. Comput Hum Behav 75:891–902. https://doi.org/10.1016/j.chb.2017.07.002
7. Bock BC, Lantini R, Thind H, Walaska K, Rosen RK, Fava JL, Barnett NP, Scott-Sheldon LA (2016) The mobile phone affinity scale: enhancement and refinement. JMIR mHealth and uHealth 4(4):e134. https://doi.org/10.2196/mhealth.6705

8. Kim SM, Huh HJ, Cho H, Kwon M, Choi JH, Ahn HJ, Lee SW, Kim YJ, Kim DJ (2014) The effect of depression, impulsivity, and resilience on smartphone addiction in university students. J Korean Neuropsychiatr Assoc 53(4):214–220. https://doi.org/10.4306/jknpa.2014.53.4.214
9. Lin Y-H, Chang L-R, Lee Y-H, Tseng H-W, Kuo TB, Chen S-H (2014) Development and validation of the smartphone addiction inventory (SPAI). PloS One 9(6):e98312. https://doi.org/10.1371/journal.pone.0098312
10. Arpaci I (2021) Relationships between early maladaptive schemas and smartphone addiction: the moderating role of mindfulness. Int J Ment Heal Addict. https://doi.org/10.1007/s11469-019-00186-y
11. Arpaci I, Kocadag Unver T (2020) Moderating role of gender in the relationship between big five personality traits and smartphone addiction. Psychiatr Q 91(2):577–585. https://doi.org/10.1007/s11126-020-09718-5

Diabetes Mellitus Prediction and Feature Importance Score Finding Using Extreme Gradient Boosting

Laboni Akter and Ferdib-Al-Islam

Abstract Day by day, diabetes mellitus has become a typical malady to humankind from youthful to old people. The development of diabetic patients is expanding step by step because of different causes, for example, bacterial or viral disease, harmful or compound substance blend in with the food, immune system response, corpulence, terrible eating regimen, change in ways of life, dietary patterns, natural contamination, and so on. Consequently, diagnosing diabetes is extremely fundamental to spare human life from diabetes. In healthcare services, this scientific procedure is completed utilizing machine learning techniques for decomposing clinical information to build predictive models to do clinical discoveries. This paper presents a diabetes prediction framework for finding diabetes in its beginning phase from patients' different clinical symptoms. Besides, this paper investigates the significant features in diabetes prediction using Extreme Gradient Boosting (XGBoost) algorithm. It has been shown that the proposed system can predict diabetes more perfectly from patients' clinical data on different performance metrics (accuracy was 94.23%, sensitivity was 94.74%, and specificity was 93.94%) with feature important scores.

Keywords XGBoost algorithm · Diabetes mellitus · Exploratory data analysis · Correlation · Feature importance score

1 Introduction

Diabetes is one of the normal and energetically expanding illness. It is discovered that diabetes is an extremely serious issue in the majority of nations. Diabetes is a condition wherein your body can't deliver the necessary measure of insulin which is expected to manage the measure of sugar in the body [1]. This may prompt different maladies like coronary illness, kidney infection, visual deficiency, nerve harm, and so forth. In 2011, nearly 8.4 million or 10% of the absolute common people have

L. Akter · Ferdib-Al-Islam (✉)
Khulna University of Engineering & Technology, Khulna 9203, Bangladesh

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_52

diabetes, and as per the "International Diabetes Federation (IDF)" in Bangladesh and the pervasiveness of diabetes mellitus among grown-ups will be expanded to 13% by 2030 [2]. Diabetes mellitus has a quick banner of high glucose, along with a couple of symptoms including constant pee, extended thirst, extended longing, and weight decrease. Patients with diabetes generally require steady treatment, else, it will conceivably incite various risky perilous complexities [3].

In this cutting edge period of innovation, computational intelligence can assist us with detecting illnesses precisely and can set aside time and cash. Machine Learning is a significant area of Computer Science that stays utilized for prediction. It focuses the way toward finding novel data from past known data. To predict a malady utilizing machine learning draws near, we need its manifestations alongside clinical data [4]. Indications are a significant cause for fresh patients and beginning phase expectations since they have no statistics aside from manifestations. We additionally need clinical information for analysis and finding novel outcomes from data [5].

In this paper, a model for diabetes prediction to detect diabetes in its early phase has been implemented and also shown the most significant features for developing diabetes in patients using the extreme gradient boosting (XGBoost) algorithm.

The leftover parts of this paper have been arranged as follows—Sect. 2 describes the recent researches in diabetes prediction applying machine learning algorithms. The implementation of this work has been represented in the Sect. 3. The results have been analyzed in the Sect. 4. The Sect. 5 presents the paper concludes with future works.

2 Literature Review

Singh et al. [6] introduced a diabetes forecast framework for diabetes conclusion. Dataset was gathered from the "University of California, Irvine (UCI)" dataset repository. This dataset contains a clinical report of 768 people and 8 features. Naïve Bayes (NB), function-oriented multilayer perceptron (MLP), decision tree-based random forests (RF) were utilized to construct the machine learning model. Furthermore, 10-fold cross-validation was applied in the dataset to assess the exhibition of the machine learning models. Srinivas et al. [7] created the utilization of data mining procedures in social insurance then the anticipation of respiratory failures. This examination applied clinical profiles as, age, pulse, glucose, and sex, and estimated the likelihood of patients receiving heart and kidney issues.

Islam et al. [8] used Naive Bayes, Logistic Regression, and Decision Trees algorithms to the expectation of Diabetes at a primary phase using Data Mining methods. The dataset was acquired from Sylhet Diabetes Hospital in Sylhet, Bangladesh. The number of attributes and instances were 16 and 520. Accuracy were 88% and 91%. For Naive Bayes, Logistic Regression, and Decision Trees, Precisions were 0.879, 0.924, 0.957, Recalls were 0.874, 0.924, 0.956, and F-measures were 0.875, 0.924, and 0.956. The main limitation of this article, there

was no feature importance calculation. Hashi et al. [9] had proposed for foreseeing diabetes utilizing the data mining procedures. The dataset was gathered from the Pima Indians Diabetes Database of the "National Institute of Diabetes and Digestive and Kidney Diseases". The framework has been created utilizing dual notable calculations that were Decision Tree calculation (C4.5) and KNN algorithm. There were 768 samples with 8 numerical attributes where 500 were negative and 268 are positive occurrences. The accuracy for C4.5 and KNN was 76.96% and 90.43% separately.

Durairaj and Ranjani [10] examined the expected utilization of classification algorithms, for example, rule-based techniques, decision tree calculation, Naïve Bayes, and artificial neural system (ANN) to the gigantic capacity of medicinal services information. In this article, clinical issues have been broken down and assessed, for example, coronary illness and circulatory pressure. Mercaldo et al. [11] built up a model for the expectation to segregate among diabetes-influenced people and not-influenced ones utilizing Machine Learning Algorithm. The classification algorithms were J48, Hoeffding Tree, Multilayer Perceptron, Bayes Net, and Random Forest. The precision was 0.757 and recall was 0.762. The dataset contains 8 features and 768 instances cases and entirely patients in the dataset were women in any event 21 years of age. Sa'di et al. [12] analyzed type-2 diabetes, this paper has matched the general classification accuracy of Naive Bayes (accuracy was 76.95%), RBF Network (accuracy was 74.35%), and J48 (accuracy was 76.52%) algorithms.

3 Methodology

The methodology of the proposed work has been divided into the following steps:

- Data Collection and Preprocessing
- Exploratory Data Analysis
- Extreme Gradient Boosting for Classification
- Feature Importance Finding.

The proposed system architecture has been illustrated in Fig. 1.

3.1 Data Collection and Preprocessing

The dataset that we have utilized, got from the "University of California, Irvine machine learning data repository" [13]. This contains 16 attributes with a target column of 520 instances. This dataset has been made from an immediate survey of individuals who had diabetes mellitus, and who were still non-diabetic yet having not many or more indications. The data has been gathered from the patients

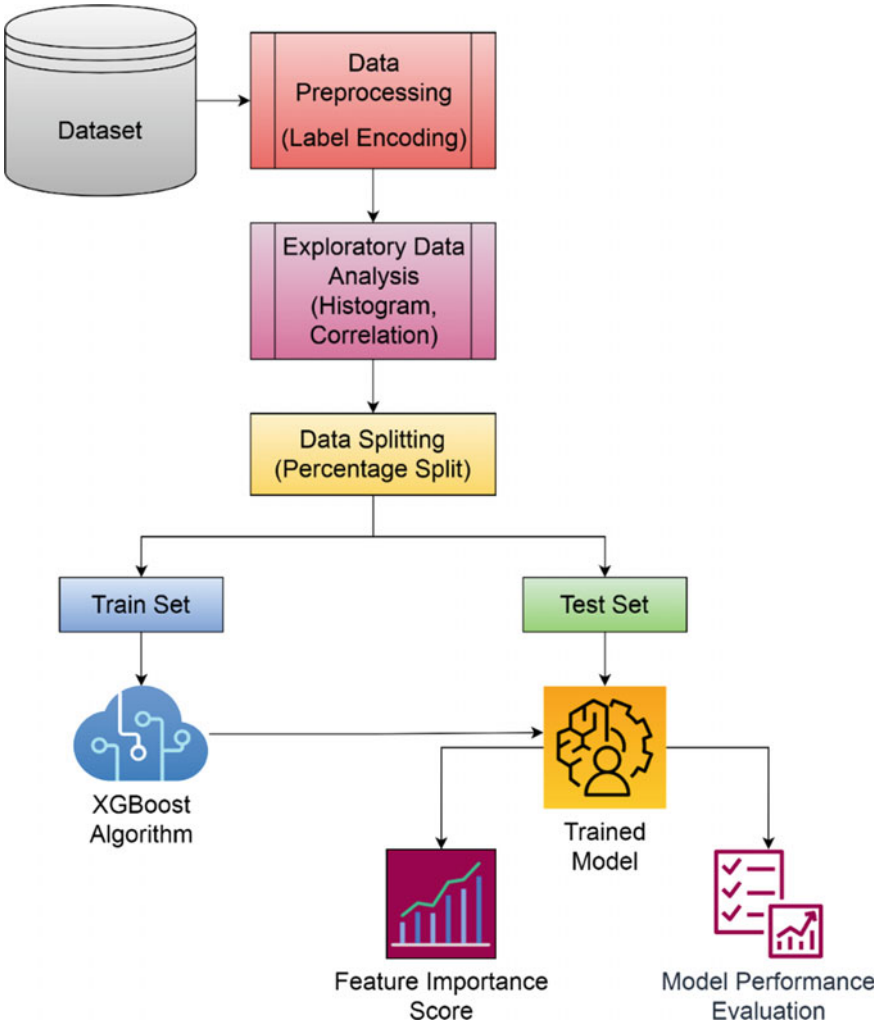


Fig. 1 Proposed system architecture

utilizing an immediate poll from the "Diabetes Hospital of Sylhet, Bangladesh" [13]. The dataset description has been described in Table 1, showing that there were no missing values. Here, "Class" was the target variable. Label Encoding alludes to changing over the labels into a numeric structure to change over it into a meaningful structure. Machine learning methods would be able to choose in a superior manner on how those labels must be worked. It is a significant preprocessing step for the organized dataset in supervised learning. So, label encoding has been performed to convert the categorical entities into integers to feed those values into the

Table 1 Dataset information

Sl. No	Column	Non-null count
0	Age	520
1	Gender	520
2	Polyuria	520
3	Polydipsia	520
4	Sudden weight loss	520
5	Weakness	520
6	Polyphagia	520
7	Genital thrush	520
8	Visual blurring	520
9	Itching	520
10	Irritability	520
11	Delayed healing	520
12	Partial paresis	520
13	Muscle stiffness	520
14	Alopecia	520
15	Obesity	520
16	Class (Target variable)	520

classification algorithm. 80% of dataused in training and 20% of data used in testing, which means, there were 416 instances in the training set and 104 instances were in the test set.

3.2 Exploratory Data Analysis

After the preprocessing step, it is needed to realize how much the correlation between the variables is. The correlation coefficient can be utilized to mirror the relationship between the variables. It is additionally founded on the scattering of the two variables and their particular midpoints. The two differences are duplicated to mirror the level of relationship between the two variables. The correlation coefficient is estimated on a scale from −1 to +1. Correlation unquestionably impacts feature significance. Implying that if the features are exceptionally correlated, there would be a significant level of excess if keeping them all. Since two features are associated that implies a change in one will change another. So there is no compelling reason to keep every one of them. Figure 2 was representing the correlation matrix of the features and it was not having any strong correlation. So, it was not needed to eliminate any feature.

A histogram is a commonly utilized data visualization method in machine learning. Histograms plot the data by partitioning it into spans called ‘bins’. It is utilized to examine the frequency distribution.

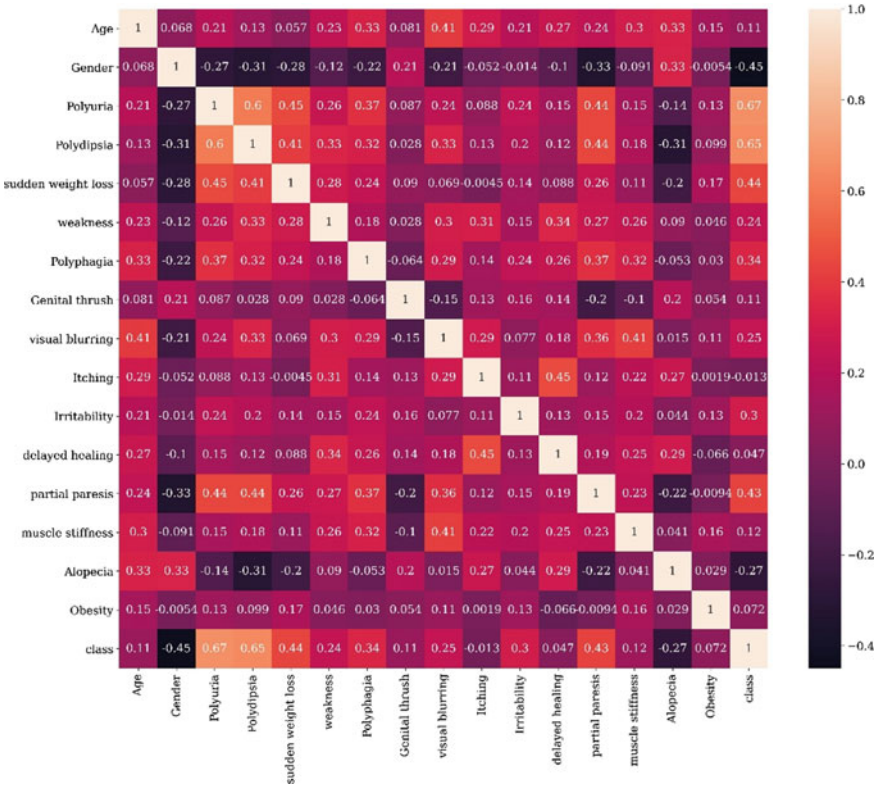


Fig. 2 Correlation among variables

According to the histogram illustrated in Fig. 3, it has been found that female persons are having diabetes mellitus more than male persons. Here, "0" means the diabetes negative cases and "1" means the positive cases, where, the red colored bin represents the females and the blue colored bin represents the males.

The age-wise class distribution has been illustrated in Fig. 4. In Fig. 4, "0" means the diabetes negative cases and "1" means the positive cases. Age distribution starts from 0 and ends up at 50. It can be seen that most of the diabetes mellitus positive cases lies among the age of 11 and 44. Deeper insight has shown that people at young ages were suffering from diabetes mellitus more than others.

3.3 Extreme Gradient Boosting for Classification

Extreme Gradient Boosting (XGBoost) is an approach that has been commanding machine learning for tabular data. XGBoost is an execution of gradient-boosted decision trees proposed for speed and execution [14]. Boosting is an ensemble

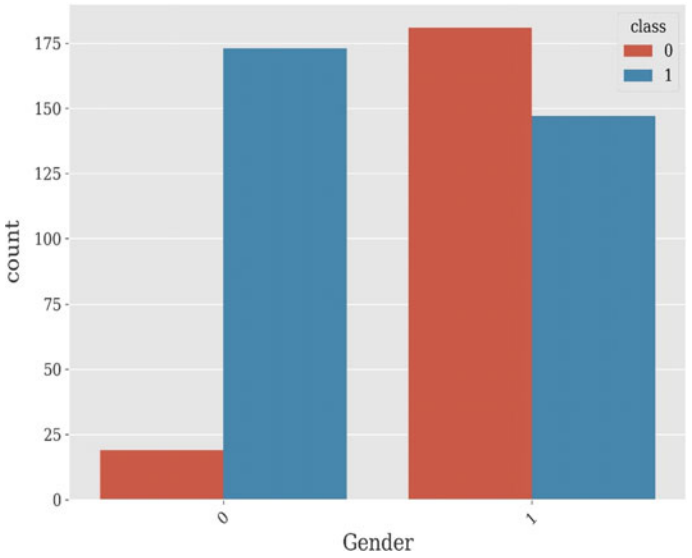


Fig. 3 Gender-wise class distribution

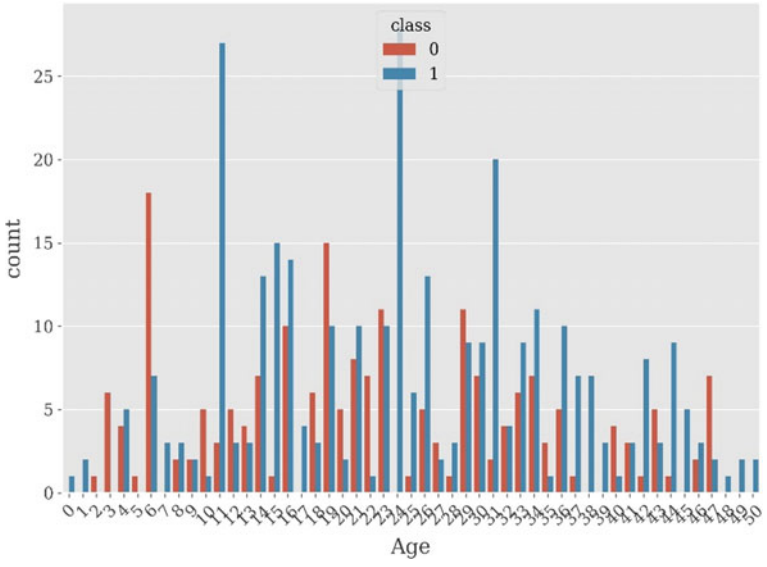


Fig. 4 Age-wise class distribution

method where novel models are included to compute the error made by current models. Models are comprised consecutively until no additional boosts can be made. A graphical representation of the XGBoost algorithm's workflow has been

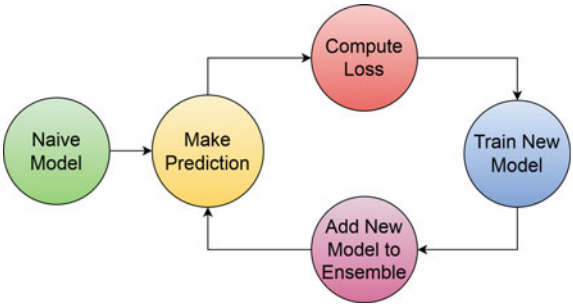


Fig. 5 XGBoost algorithm workflow

demonstrated in Fig. 5. Extreme Gradient Boosting is a procedure where novel models are made to anticipate the residuals of prior models and afterward included to make the final forecast. It is called gradient boosting because it utilizes a gradient descent method to lessen the loss while including novel models.

The objective function is an aggregate of a particular loss assessed generally classifications and a whole of regularization term for all classifiers. Mathematically, it can be represented as in (1):

$$obj(\theta) = \sum_{i}^{n} l(y_i - \hat{y}_i) + \sum_{K=1}^{K} \Omega(f_K) \tag{1}$$

XGBoost algorithm has shown significantly improved performance in biomedical applications using different clinical datasets [15, 16]. This algorithm has been utilized in this work to anticipate if the patient travailing from Diabetic Mellitus or not (binary classification). The chosen parameters of the XGBoost classifier for this classification task have been represented in Table 2.

3.4 Feature Importance Finding

Exploring the significant features of diabetes through systematic techniques can anticipate and prevent diabetes mellitus. The facility of utilizing the XGBoost is that the importance score for every feature can be acquired generally after the

Table 2 Chosen parameter values for XGBoost classifier

Parameter name	Chosen value
booster	Gbtree
colsample_bytree	0.3
learning_rate	0.1
max_depth	4
objective	binary:logistic

decision tree is made. The feature importance score quantifies the estimation of the component in the development of the tree in the model. The importance of the feature is determined by arranging each attribute in the dataset. It is ensured by improving the measure of execution measurements by each attribute split point in every tree, and the node is accountable for the weighting and number of records. At last, the outcome of each feature in all the advancement trees is weighted and added, and afterward, the mean value becomes the importance score.

4 Result and Discussion

It was mentioned earlier that, XGBoost classifier has been utilized in this work to predict diabetes mellitus in patients' and calculating feature importance scores. The implemented system performance was measured using different performance metrics—accuracy, precision, recall, f1-score, sensitivity, and specificity using these formulas:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{2}$$

$$Sensitivity = \frac{TP}{TP + FN} \tag{3}$$

$$Specificity = \frac{TN}{TN + FP} \tag{4}$$

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

$$F1{-}Score = \frac{TP}{TP + \frac{1}{2}(FP + FN)} \tag{7}$$

The overall test accuracy was 94.23%, sensitivity and specificity were 94.74% and 93.94% respectively. The detailed classification report (class-wise accuracy, precision, recall, and f1-score) has been described in Table 3.

Table 3 Class-wise classification report of XGBoost model

Class	Precision (%)	Recall (%)	F1-Score (%)	Support
Negative	90	95	92	38
Positive	97	94	95	66

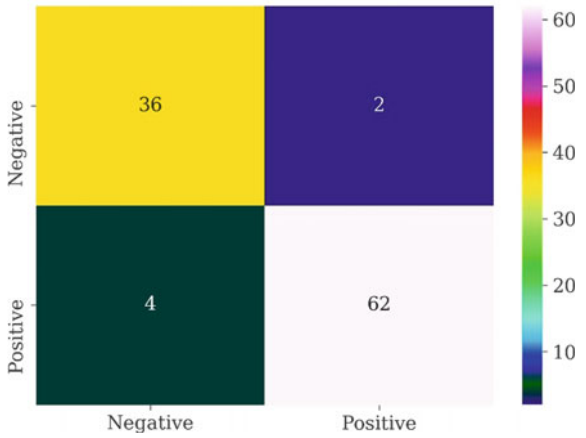


Fig. 6 Confusion matrix from XGBoost model

The confusion matrix of the XGBoost model has been illustrated in Fig. 6. There were 98 correct classifications and 6 wrong classifications. After making the diabetes mellitus prediction, the feature importance score which has been obtained by using XGBoost has been calculated, as shown in Fig. 7. "Age", "delayed_healing", and "muscle_stiffness" were the top-3 features in the list. Aging can make diabetes

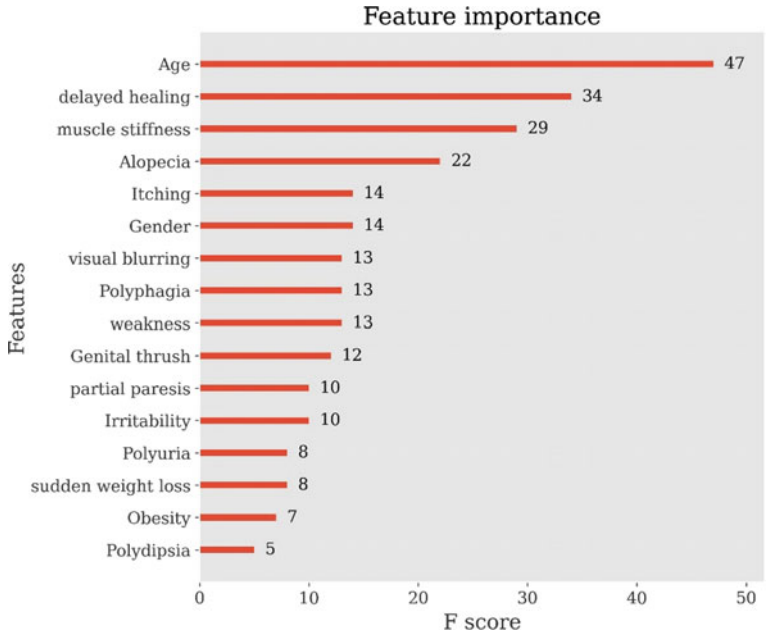


Fig. 7 Feature importance score from XGBoost model

Table 4 Proposed model comparison with others'

Author	Model used	Accuracy (%)	Precision (%)	Recall (%)	Feature importance calculation
Islam et al. [8]	Naïve Bayes	88	88.1	88	No
	Logistic Regression	91	91.1	91	
Hashi et al. [9]	KNN	90.43	N/A	N/A	No
	C4.5	76.96	N/A	N/A	
Sa'di et al. [12]	Naïve Bayes	76.95	N/A	N/A	No
	RBF	74.35	N/A	N/A	
	J48	76.52	N/A	N/A	
Li et al. [17]	RFE + XGBoost	80.2	N/A	70.6	Yes
Proposed work	XGBoost	94.23	90	95	Yes

harder to control. One purpose behind this is that insulin obstruction increments and glucose resilience diminishes with age. At the point when the glucose level is higher than typical, it keeps supplements and oxygen from stimulating cells, keeps the immune system from working productively, and builds aggravation in the body's cells. These causes delayed healing. Individuals with diabetes can encounter muscle stiffness, which can be due to high blood glucose levels as well as nerve damages.

A comparison between the proposed method and other existing works has been represented in Table 4. It has been shown that the proposed work showed better performance than the existing works in different performance metrics along with the feature importance score, which was not done in the existing works.

5 Conclusion

Diabetes Mellitus is one of the quickest developing constant perilous sicknesses that have just influenced 422 million individuals overall as per the findings of the World Health Organization (WHO), in 2018. Because of a moderately long asymptomatic stage, the initial identification of diabetes mellitus is constantly wanted for a clinically significant result. The authors implemented extreme gradient boosting (XGBoost) for both classification and finding feature importance score. The accuracy (94.23%), sensitivity (94.74%), and specificity (93.94%) of this system were satisfactory. However, the selected important features can improve the overall performance more but still, it can be used for early detection of diabetes in patients. Further research with a larger dataset and the development of an automated toolkit will facilitate both physicians and patients to diagnose diabetes rapidly.

References

1. Insulin—Diabetes. https://www.diabetes.co.uk/body/insulin.html. Last accessed 2021/01/05
2. Whiting D et al (2011) IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030. Diabetes Res Clin Pract 94(3):311–321
3. Clark N, Fox K, Grandy S (2007) Symptoms of diabetes and their association with the risk and presence of diabetes: findings from the study to help improve early evaluation and management of risk factors leading to diabetes (SHIELD). Diabetes Care 30:2868–2873. https://doi.org/10.2337/dc07-0816
4. Harris M et al (1992) Onset of NIDDM occurs at least 4–7 yr before clinical diagnosis. Diabetes Care 15(7):815–819
5. Musacchio N et al (2020) Artificial intelligence and big data in diabetes care: a position statement of the Italian association of medical diabetologists. J Med Internet Res 22(6): e16922
6. Singh D et al (2017) Diabetes prediction using medical data. J Comput Intell Bioinforma 10 (1):1–8
7. Srinivas K, Kavihta RB, Govrdhan A (2010) Applications of data mining techniques in healthcare and prediction of heart attacks. Int J Comp Sci Eng 2(2):250–255
8. Islam M et al (2019) Likelihood prediction of diabetes at early stage using data mining techniques. In: Computer vision and machine intelligence in medical image analysis, pp 113–125
9. Hashi E et al (2017) An expert clinical decision support system to predict disease using classification techniques. In: 2017 International conference on electrical, computer and communication engineering (ECCE)
10. Durairaj M, Ranjani V (2013) Data mining applications in healthcare sector: a study. Int J Sci Technol Res 2:29–35
11. Mercaldo F et al (2017) Diabetes mellitus affected patients classification and diagnosis through machine learning techniques. Procedia Comp Sci 112:2519–2528
12. Sa'di S et al (2015) Comparison of data mining algorithms in the diagnosis of type II diabetes. Int J Comput Sci Appl (IJCSA) 5(5):1–12
13. UCI machine learning repository: early stage diabetes risk prediction dataset. Data set. https://archive.ics.uci.edu/ml/datasets/Early+stage+diabetes+risk+prediction+dataset. Last accessed 2021/01/05
14. Kadiyala A, Kumar A (2018) Applications of python to evaluate the performance of decision tree-based boosting algorithms. Environ Prog Sustainable Energy 37(2):618–623
15. Ma B et al (2020) Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Comput Biol Med 121:103761
16. Zou D (2014) The research on the boosting decision tree algorithm for intelligent medical system. Appl Mech Mater 539:365–368
17. Li M et al (2020) Diabetes prediction based on XGBoost algorithm. IOP Conf Ser: Mater Sci Eng 768:072093

Factors Influencing E-payment Adoption. Case Study: Pandemic COVID-19

Dounia Dahab and Fatine Bouqlila

Abstract The context of the COVID-19 pandemic has driven consumers to use remote and contactless services. Indeed, and in order to avoid any kind of contamination, it was recommended by international health authorities to adopt a policy of physical distance. This policy has had an impact on consumer behavior and on the way in which consumers make a purchase. Thus, some consumers, fearing the risk of being contaminated, have begun to prefer online purchases and consequently the use of electronic payments (E-payments). This study aims to analyze the intention to adopt E-payments during the COVID-19 pandemic. It suggests the effect of perceived risk, perceived security, trust, social distancing, and Internet experience on the behavioral intent to adopt E-payments in Morocco during the pandemic. The proposed research model was developed by extending the Technology Acceptance Model (TAM) to explain and predict consumers' intentions to adopt these e-services.

This study stands out for its originality in the context of the COVID-19 pandemic, new variables such as COVID-19 perceived risk and social distancing with multiple impacts on consumer purchasing behavior are included in the model. It thus enriches the existing literature on E-payments adoption models, especially since there is no similar research on the Moroccan market. In addition, the conceptual model proposed in this study can provide important guidance to firms wishing to define appropriate strategies to promote the use of electronic payments.

Keywords Perceived COVID-19 risk · Perceived security · Trust · Social distancing · Internet experience · Perceived usefulness · E-payments adoption

D. Dahab · F. Bouqlila (✉)
ISCAE Casablanca, Route de Nouasseur, 8114 Casablanca, Morocco
e-mail: ddahab@groupeiscae.ma

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_53

1 Introduction

The COVID-19 pandemic has affected several sectors including industry, education, tourism and other sectors in various countries [1, 2]. Indeed, publications have shown that the significant impacts of this pandemic on, for example, the socio-economy [3, 4] and finance [5] are mainly due to the policy of imposed social distancing.

This policy of social distancing involving the adoption of physical distance between individuals has significantly reduced the services provided to individuals [2] and affected the way in which the consumer makes a payment [6]. Some people fear the risk of using physical money regarding the risk of contamination of coins and bills by the virus. As a result, several companies have reviewed the way to offer their services to their customers and start offering remote services. Those services have been well received by a certain category of customers who have an appetite for and ease of consuming these categories of services.

Among the services offered remotely, the "Untact Service". Untact Service refers to service that is provided without face-to-face encounters between employees and customers through the use of digital technologies [7]. Thus, thanks to the development of Industry 4.0, various players have started to offer these services. They are spreading in various areas of daily life such as mobile and Internet banking services and web and mobile ordering and purchasing services, including e-commerce [8]. The consumption of these e-services implies the realization of electronic payments.

The latter have also experienced a major craze during the pandemic as attested by the CMI (Centre Monétique Interbancaire)[1], growth of online payments activity by Moroccan card of +41.4% in number of transactions, rising from 6.8 million transactions during the first half of 2019 to 9.7 million transactions during the first 9 months of 2020) [9]. This rebound in transactions can be explained by the fact that E-payments allow the customer to carry out a transaction without physical contact with the merchant and that they also promote the manipulation of electronic money as recommended by the OMS [10] at the expense of cash. The latter can be the carrier of the virus when used by an infected person.

Our study, which pays particular attention to the issue of the COVID-19 pandemic and its relationship with payment behavior, attempts to explain the factors that may impact the intention to adopt E-payments during the pandemic.

Previous research on the use of E-payments in general has yielded a variety of results. Pei et al. [11] research has shown that perceived benefit and trust are the key factors determining user adoption of E-payment tools. Barkhordari et al. [12] research on the determinants of adoption of Electronic Payment Systems (EPS) has shown that perceived security and trust have a positive impact on the adoption of these systems. The Internet experience is also an important driver for consumer adoption of electronic and mobile payments [13]. Social distancing is also a relevant variable in the context of the pandemic that could impact E-payments adoption

[1] CMI is a financial institution providing Electronic payment solutions in Morocco.

as it influences consumer purchasing behavior [14]. The perceived risk has also been proven to have a significant impact on E-payments adoption. Indeed, previous research on the adoption of online banking services concluded that perceived risk impacts the consumer's intention to use these e-services [15]. In general, most studies that have addressed this factor have shown that the effect of perceived risk on behavior is negative. However, in the context of the pandemic, the conclusion might be different because the risk is associated with infection with a deadly virus. As mentioned above, there is a possibility of transmission of the virus through contact with a sick person or through physical currency. Thus, the perceived risk associated with the transmission of SARS-Cov2 could have a positive impact on consumers' intention to use E-payments. Unfortunately, research on this element remains poor. Although many researchers have studied the factors influencing the adoption of E-payments, studies examining the effect of the COVID-19 pandemic on the way consumers make a payment remain scarce. Because of this finding, our current study contributes to the literature by presenting a model including perceived COVID-19 risk, social distancing, trust, perceived security and Internet experience on consumers' intention to use E-payments during the COVID-19 epidemic.

Our paper is organized as follows. First, we will discuss our literature review, then we will present our conceptual model, and finally we will discuss future research and the conclusion of our work.

2 Literature Review

In this section, we present a review of the literature, defining E-payments, and discussing the TAM.

2.1 E-payments

Electronic payments refer to online transactions made via the Internet [16]. They have many advantages [17]. They are easy, convenient and allow transferring money anywhere in the world. They also make it possible to keep written records and are less risky. However, they also have some drawbacks, such as the difficulty of use for the uninitiated and the risk of data leaks.

After presenting the obstacles and advantages related to the use of E-payments, knowledge of the different factors influencing their adoption from the consumer's point of view during the crisis is important to promote their use. Indeed, their success depends mainly on their adoption, especially since before the crisis; the use of E-payments in Morocco was low. According to the CMI, 91% of cards are used exclusively for ATM withdrawals, and only 9% for electronic payments [18]. The research then addresses this gap by studying the adoption of E-payments through a rigorous adoption model which is TAM.

2.2 TAM (Technology Acceptance Model)

In this research, we use the TAM model as the theoretical basis as it is considered the most robust and rigorous model for explaining and predicting technology intention and use [19, 20]. TAM has been applied in many areas. For example, Bailey et al. [21] have used it to explore the adoption of contactless payment by US consumers. Liébana-Cabanillas et al. [22] borrowed it to determine the factors that affect the acceptance of SMS and NFC mobile payment systems.

Our study adopts the final version of the TAM [23] and integrates variables adapted to the study of E-payments, which are perceived risk and security, trust, social distancing, and Internet experience (which is an important variable facilitating website navigation and making payment). Most of these variables influence the adoption of E-payments [6, 12].

However, the majority of research using the TAM has focused mainly on three factors, which are trust, security and perceived risk. Indeed, Aji et al. [6] investigated the influence of 3 variables (Perceived usefulness, risk and government support) on the intention to adopt e-wallet during the COVID-19 pandemic. Their results showed that only perceived risk and usefulness had an influence on the adoption of these e-services. Salloum et al. [24] extended the TAM with a single variable, trust, to examine students' intention to use EPS. Their results suggested that trust is a significant factor in predicting perceived usefulness, ease of use and adoption of E-payments. Barkhordari et al. [12] research focused on studying the important factors that influence trust in EPS in Iran. The results showed that perceived security and trust have a positive impact on the use of EPS.

The originality of this research is the introduction of social distancing, perceived COVID-19 risk as independent variables which might explain intention to adopt E-payments. Therefore, previous studies have not used these factors in a single setting and have not studied them in the context of the pandemic. Finally, most of the studies dealing with this topic were conducted outside Morocco. Following this, it is interesting to examine these factors, in the Moroccan context and during the pandemic period, to better understand the factors impacting the E-payments decision during this period.

In the following section, we will present the links between the variables used to design our conceptual model of E-payments adoption.

3 Conceptual Model

3.1 Impact of COVID-19 on Consumer Behavior

The new CORONAVIRIUS has innumerable negative impacts on people's health, and on the economy and finances [5]. It also impacts consumer behavior as noticed by Hollingsworth et al. [25] which show that purchasing behavior is influenced by

diseases caused by epidemics. Studies have shown that the use of online retail applications and mobile e-commerce applications by inactive and new users increased significantly during this period [26]. Indeed, consumers have turned online to find the products they need. In France, for example, the Observatoire de la Sécurité des Moyens de Paiement (OSMP) [27] has noted an acceleration in the use of dematerialized means of payment, with a spectacular increase in contactless, to the detriment of checks and cash in particular. In Morocco, during the pandemic, the online payment activity of Moroccan cards increased by +28.1% in amount, during the first 9 months of 2020 [9].

The policy of social distancing and quarantine in case of contamination has pushed people to stay at home. Consumers are starting to carry out most of their activity without being in contact with others. Moreover, a study conducted by Le Desk [28] states that new spending behaviors emerged during the pandemic period, where a category of Moroccan consumers gradually turned to online shopping. According to the same study, only 35% of Moroccan consumers think they are back to their normal pre-pandemic habits and that 26% would opt for online purchase of most of their products.

3.2 Perceived Risk and Perceived Usefulness

In the past, perceived risk was mainly related to fraud and product quality, and now it covers all types of risks: financial, product performance, social, psychological, physical or temporal when users conduct online transactions [29]. Mäser et al. [30] have added another dimension of risk, which is the risk of illness, a relevant facet of risk in the context of our study.

Thus, in our research, we retain the definition of Aji et al. [6] which considers the perceived risk as the situation in which consumers are unsure of the presence of traces of coronavirus on physical money. From there, we consider that the dimension of risk addressed is more related to the risk of disease because customers fear being infected by exchanging physical money [31]. E-payments are, therefore, an alternative to avoid the risk of COVID contamination. As a result, and given that consumers perceive advantages related to the use of E-payments, they will be more likely to adopt them [32].

This link between perceived risk and perceived usefulness has been proven in the literature. It has been demonstrated that perceived risk has a significant effect on perceived usefulness [33]. The Hampshire [34] research on the influence of trust and risk on British consumer's payment showed that perceived risk negatively influences perceived usefulness. Since in this study perceived risk refers to COVID-19 Risk, it is supposed to have positive effects on perceived usefulness, in contrast to previous studies where negative effect of risk on perceived usefulness and intention to adopt a technology was demonstrated. Therefore, we hypothesize as follows:

H1: Perceived COVID-19 risk positively affects perceived usefulness of E-payments.

3.3 Social Distancing and Perceived Usefulness

Social distancing refers to the reduction of interactions between individuals [35]. This variable has been addressed very regularly in the current COVID context. It is one of the most recommended policies in the world to reduce the risk of spread of the SARS-Cov2 virus during the pandemic [36]. Many countries have taken unprecedented measures to prevent social contact and slow the spread of the virus. These measures can all be described as "social distancing", and are particularly effective for diseases (such as COVID-19) that are transmitted by respiratory droplets, which are highly transmissible, in the case of close contact with sick people [37].

Social distancing has led to a change in the consumption habits of a significant portion of the population. Indeed, during the pandemic, some consumers turned to online shopping and home delivery to minimize contact. Also, avoidance of social contact has also had an impact on the number and types of out-of-home activities that people engage in [35]. Finally, there was an increase in the use of online retail applications and mobile e-commerce applications by inactive and new users during this period [26].

Based on these elements, we can see that social distancing impacts consumer behavior and prompts them to prioritize online consumption in the context of the pandemic. It could, then, be a determining factor in the adoption of E-payments in this context. Since social distancing measures have obvious direct positive effects on health, as long as they prevent the contamination of people by the virus [35], it is likely that these measures may reflect the usefulness of E-payments which ban the use of cash. This is why we make the following hypothesis:

H2: Social distancing positively affects the perceived usefulness of E-payments.

3.4 Perceived Security and Perceived Usefulness

Security is a major concern as customers are reluctant to enter personal information online. In the area of E-payments, it represents how EPS can protect consumer transactions [38]. Indeed, the consumer's perception of security has a direct impact on the use of E-payments because of the sensitivity of the services offered by these systems [12].

Thus, security could represent a determining factor in the consumer's adoption of E-payments. Research has indicated that if consumers perceive more online

security cues, their intentions to use E-payments increase even if the systems or services are perceived as unconvincing [39].

In our study, we will focus on the impact of perceived security on the usefulness of E-payments. Indeed, if the consumer perceives the system to be secure, he perceives its usefulness and is encouraged to use it. This has been proven in the research conducted by Atarodi et al. [40] which has shown a significant impact of perceived security on perceived usefulness in the adoption of technologies. Also, Basak et al. [41] research on intention to adopt e-commerce has demonstrated a significant influence of security on perceived usefulness and perceived ease of use. To confirm the effect of security on the perceived usefulness of E-payments, the following hypothesis is proposed:

H3: Perceived security positively affects perceived usefulness of E-payments.

3.5 Trust and Perceived Usefulness

In the EPS context, trust is defined as the belief of consumers that E-payments transactions will be processed according to their expectations [42]. With the evolution of payment methods, it has become a very important factor for the success of mobile payment systems [43] and EPS in general. Moreover, several studies have proven that trust has a positive impact on the adoption of EPS [12, 44]. It also influences other factors such as perceived usefulness, ease of use, and risk, which in turn influence the adoption of mobile payment systems [43], a category of EPS.

From these elements, we can assume that trust has an impact on personal beliefs about perceived usefulness, which in turn has an impact on the behavioral intention to use E-payments. This is explained by the fact that once this trust is established, users are more likely to have more confidence in the perceived usefulness of E-payments. For this reason, we retain the following hypothesis:

H4: Trust positively affects the perceived usefulness of E-payments.

3.6 Internet Experience and Perceived Usefulness

In our study, we will use Rose's et al. [45] definition of Internet experience which states that the experience remains in the customer's memory and helps them shape future decisions based on their daily learning. In the context of online shopping, Naseri's et al. [46] research showed that the likelihood of making online purchases was significantly higher among consumers who used the Internet, e-mail or chat rooms more frequently and who had experience with online banking, trading and administration. Also, Hernández et al. [47] confirm that prior experience with the Internet is a necessary condition for the adoption of e-commerce.

From these elements, we can see that having prior experience with the Internet significantly reduces the time and effort involved in both learning and conducting online operations, which could promote the use of E-payments. In our study, and in line with the above, we believe that the more experienced users are in using the Internet, in conducting their online transactions, the more likely they are to adopt E-payments. Also, the Internet experience positively influences the perceived usefulness of using mobile payment [13]. Thus, in our research, we will assume that the Internet experience positively influences the perceived usefulness of E-payments. We then put forward the following hypothesis:

H5: Internet experience positively affects the perceived usefulness of E-payments.

3.7 Perceived Usefulness and Intention to Adopt E-payments

Perceived usefulness has been defined as the belief that the use of a particular system would improve job performance [19]. On several occasions, perceived usefulness has been regarded as a perceived benefit, which is why Roger [48] sees it as the extent to which an innovation is perceived to be better than its predecessor.

Perceived usefulness strongly influences consumer behavior. It has been shown to have a positive impact on the intention to adopt the Internet, mobile services [49] and mobile banking [50, 51]. In the field of E-payments, previous researches consider that perceived benefit is an important factor in the acceptance of EPS [52]. Therefore, the study makes following hypothesis:

H6: The perceived usefulness of E-payments has a positive impact on the intention of its adoption.

In consideration of the discussion above, the proposed model for E-payments adoption is shown in Fig. 1.

4 Discussion and Future Research

This study contributes to the current research stream on consumer behavior during COVID-19 pandemic and the literature on user acceptance of E-payments. The major limitation of this research is that the model proposed hasn't been confronted to the field. In order to validate the model presented in this research, future research will focus on the verification and testing of the previously defined hypotheses using the measurement scales already identified in the literature (See Appendix).

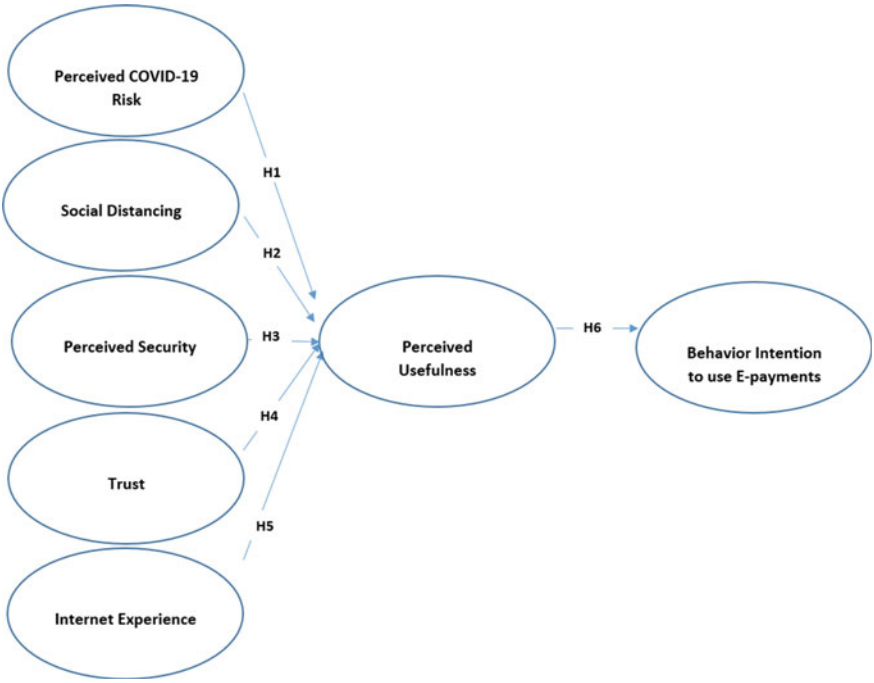


Fig. 1 E-payments adoption model

A self-administered questionnaire along with demographic information will be sent to respondents through online mode. The questionnaire is composed of two parts. The first part includes questions that operationalize the variables in our model. The second part contains respondent demographics data such as gender, age, occupation and income. The target population of the study will be consumers with bank accounts and credit cards, since in Morocco this is the most used method for electronic payments.

We will proceed, then, by validating the scales, and analyzing the data after the selection of the statistical methods to be used. For the choice of the statistical methods, we will use two methods which will be carried out in two stages: First, we will use exploratory factorial analysis to purge the measurement scales from a pre-test sample. We will use PCA: Principal Component Analysis. In a second step, we will use structural equation methods under the PLS approach to analyze the final sample. Indeed, the PLS approach has been considered the most popular method in a wide variety of scientific communities [53] for several years. It allows to model complex relationships between observed and latent variables.

5 Conclusion

The model proposed in this study will be used to examine consumer intention to use E-payments in Morocco during the COVID-19 pandemic. This model uses TAM to predict consumer intent to use E-payments technology in Morocco. Five external variables were incorporated to the TAM. These variables are: perceived COVID-19 risk and security, trust, social distancing, and Internet experience. For the social distancing variable, we believe that this variable can be studied beyond the current COVID-19 pandemic. Indeed, it is now becoming an essential factor to be taken into account in our consumption habits as subsequent waves of this virus [54] or new pandemics may occur in the future. Thus, consumers may increasingly fear social contact even if the rules of physical distance are no longer in force. Also, some people feel uncomfortable or embarrassed in their human relationships [55, 56] and prefer to avoid social contact.

Also, in order to test the hypotheses of our research, we have proposed measurement scales that can be borrowed to operationalize the factors influencing consumers' intention to adopt E-payments, and serve as the basis for the empirical study. As well, the model variables could be used in future studies related to technology adoption.

Lastly, we can notice that in their attempt to encourage E-payments, Moroccan authorities haven't succeeded before COVID-19 Pandemic encouraged a great number of consumers to switch to this payment method. Our belief is that this behavior will continue after the pandemic and people may change the way they pay goods and services in Morocco. Further research need to focus on the continuance intention to use Electronic Payments after the pandemic and how trust in this payment method and satisfaction can stimulate the transition to Mobile Payment.

Appendix

See Table 1.

Table 1 Overview of questionnaire items

Construct	Source	Measurement
Perceived COVID-19 risk	Aji et al. [6]	PR 1: I am worried to get infected by coronavirus when using physical cash PR 2: I am not comfortable making payment using physical cash PR 3: I am afraid to get infected by coronavirus when using physical cash PR 4: I am worried there is a coronavirus droplet in physical cash

(continued)

Table 1 (continued)

Construct	Source	Measurement
Perceived security	Barkhordari et al. [12]	PS 1: I perceive E-payments as secure PS 2: I think the information related to user and the history of his/her payments are secure in E-payments PS 3: My previously registered information in E-payments is utilized for security transactions PS 4: I do not worry of malicious invasions into E-payments
Trust	Barkhordari et al. [12]	T1: In either roles of seller and buyer, I trust the E-payments T2: I trust the security mechanisms of E-payments T3: I trust E-payment services T4: I trust the information displayed during an E-payments process
Social distancing	Xie et al. [35]	SD1: Avoid going out of any activity due to COVID-19 SD2: Avoid contact with individuals who have influenza SD3: Avoid travelling within or between cities/local regions SD4: Avoid using public transport due to COVID-19 SD5: Avoid going to crowded places due to COVID-19
Perceived usefulness	Pang et al. [57]	I believe: PU1: E-payments can save much time and sped up money transfer during online purchasing PU2: E-payments can make online purchasing process more efficient PU3: E-payments is more useful and compared with other payment methods, all things considered
Internet experience	Su et al. [13]	IE1: I use a computer shop online often IE2: I use a computer to look at the management of my bank account often IE3: I use online banking to complete a transaction often
Behavior intention	Sivathanu [17]	BI1: I am deliberate to make payments with use of E-payments BI2: I am deliberate do all my financial transaction using the E-payments BI3: I am interested in E-payments BI4: I am deliberate to manage my accounts with the help of E-payments BI5: I am deliberate to make payment transfers with the help of E-payments BI6: I am inquisitive for about E-payments

References

1. Murphy MPA (2020) COVID-19 and emergency eLearning: consequences of the securitization of higher education for post-pandemic pedagogy. Contemp Security Policy 41(3):492–505
2. Javaid M, Haleem A, Vaishya R, Bahl S, Suman R, Vaish A (2020) Industry 4.0 technologies and their applications in fighting COVID-19 pandemic. Diab Metab Syndrome: Clin Res Rev 14(4):419–422
3. Fernandes N (2020) Economic effects of coronavirus outbreak (COVID-19) on the world economy. SSRN Electron J
4. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Losifidis C, Agha M, Agha R (2020) The socio-economic implications of the coronavirus and COVID-19 pandemic: a review—ScienceDirect. Int J Surg 78:185–193
5. Goodell JW (2020) COVID-19 and finance: agendas for future research. Fin Res Lett 35
6. Aji HM, Berakon I, Md Husin M (2020) COVID-19 and e-wallet usage intention: a multigroup analysis between Indonesia and Malaysia. Cogent Bus Manage 7(1):1–16
7. Bordoloi S, Fitzsimmons J, Fitzsimmons M (2018) Service management: operations, strategy, information technology, 9th edn. McGraw-Hill, New York
8. Lee SM, Lee DH (2020) "Untact": a new customer service strategy in the digital age. Service Bus 14(1)
9. CMI (2020) ACTIVITÉ MONÉTIQUE AU 30 SEPTEMBRE 2020. Rapport CMI Septembre 2020. File:///C:/Users/Admin/Downloads/Cmi_activite_monetique_marocaine_au_30_septembre_2020%20(1).Pdf
10. Brown D (2020) Can cash carry coronavirus? World Health Organization says use digital payments when possible. USA Today
11. Pei Y, Wang S, Fan J, Zhang M (2015) An empirical study on the impact of perceived benefit, risk and trust on E-payment adoption: comparing quick pay and union pay in China. In: Proceedings—2015 7th international conference on intelligent human-machine systems and cybernetics, IHMSC 2015, vol 2, pp 198–202
12. Barkhordari M, Nourollah Z, Mashayekhi H, Mashayekhi Y, Ahangar MS (2017) Factors influencing adoption of e-payment systems: an empirical study on Iranian customers. ISeB 15 (1):89–116
13. Su P, Wang L, Yan J (2017) How users' internet experience affects the adoption of mobile payment: a mediation model. Technol Anal Strat Manage 30(2):186–197. https://doi.org/10.1080/09537325.2017.1297788
14. Shi K, De Vos J, Yang Y, Witlox F (2019) Does e-shopping replace shopping trips? Empirical evidence from Chengdu, China. Transp Res Part A: Policy Practice 122:21–33
15. Marafon DL, Basso K, Espartel LB, de Barcellos MD, Rech E (2018) Perceived risk and intention to use internet banking: the effects of self-confidence and risk acceptance. Int J Bank Mark 36(2):277–289
16. Armesh H, Saljoughi Z, Kord B (2010) Electronic payment and its implications. J Contemp Res Bus:246–256
17. Sivathanu B (2019) Adoption of digital payment systems in the era of demonetization in India: an empirical study. J Sci Technol Policy Manage 10(1):143–171
18. LeBoursier (2019) Http://Leboursier.Ma/Actus/4391/2019/04/11/Forum.-Le-PaiementElectronique-Ne-Decolle-Toujours-Pas-Au-Maroc.Html
19. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13(3):319–340
20. Marangunić N, Granić A (2015) Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 14(1):81–95
21. Bailey A, Pentina I, Mishra AS, Ben Mimoun MS (2017) Mobile payments adoption by US consumers: an extended TAM. Int J Retail Distrib Manage 45(6):626–640. https://doi.org/10.1108/IJRDM-08-2016-0144

22. Liébana-Cabanillas F, De Luna IR, Montoro-Ríosa F (2017) Intention to use new mobile payment systems: a comparative analysis of SMS and NFC payments. Econ Res-Ekonomska Istrazivanja 30(1):892–910
23. Davis, Fred D, Venkatesh V (1996) A critical assessment of potential measurement biases in the technology acceptance model: three experiments. Int J Hum Comput Stud 45(1):19–45
24. Salloum SA, Al-Emran M (2018) Factors affecting the adoption of e-payment systems by university students: extending the tam with trust. Int J Electron Bus 14(4):371–389
25. Hollingsworth TD, Ferguson NM, Anderson RM (2007) Frequent travelers and rate of spread of epidemics. Emerg Infect Dis 13(9):1288–1294
26. Sahli F (2020) Coronavirus (Covid19) : Quel impact sur le comportement d'achat en ligne des consommateurs tunisiens ? 2(3) :6–36
27. OSMP (2020) L'Observatoire de la sécurité des moyens de paiement. Banque de France. File:///C:/Users/Admin/Desktop/AIM/Nouveaux%20documents/Principaux%20articles/820124_osmp2019_web_vf.Pdf
28. Le Desk. https://Ledesk.Ma/2020/08/05/Post-Covid-Les-Marocains-Envisagent-de-Changer-Leurs-Habitudes-de-Consommation
29. Forsythe SM, Shi B (2003) Consumer patronage and risk perceptions in internet shopping. J Bus Res 56(11):867–875
30. Mäser B, Weiermair K (1998) Travel decision-making: from the vantage point of perceived risk and information preferences. J Travel Tour Mark 7(4):107–121. https://doi.org/10.1300/J073v07n04_06
31. Oh SH, Paek HJ, Hove T (2015) Cognitive and emotional dimensions of perceived risk characteristics, genre-specific media effects, and risk perceptions: the case of H1N1 influenza in South Korea. Asian J Commun 25(1):14–32
32. Riquelme HE, Rios RE (2010) The moderating effect of gender in the adoption of mobile banking. Int J Bank Market 28(5):328–341
33. Lee SY, Park J (2016) A study on the intention of the use of mobile payment services: application of the technology acceptance model. Korean Manage Sci Rev 33(2):65–74. https://doi.org/10.7737/kmsr.2016.33.2.065
34. Hampshire C (2017) A mixed methods empirical exploration of UK consumer perceptions of trust, risk and usefulness of mobile payments. Int J Bank Mark 35(3):354–369. https://doi.org/10.1108/IJBM-08-2016-0105
35. De Vos J (2020) The effect of COVID-19 and subsequent social distancing on travel behavior. Transp Res Interdiscip Persp 5:100121
36. Xie K, Liang B, Dulebenets MA, Mei Y (2020) The impact of risk perception on social distancing during the COVID-19 pandemic in China. Int J Environ Res Public Health 17 (17):1–17. https://doi.org/10.3390/ijerph17176256
37. Wilder-Smith A, Freedman DO (2020) Isolation, quarantine, social distancing and community containment: pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. J Travel Med 27(2). https://doi.org/10.1093/jtm/taaa020
38. Harris H, Guru BK, Avvari MV (2011) Evidence of firms' perceptions toward electronic payment systems (EPS) in Malaysia. Int J Bus Inf 6(2):226–245. Retrieved from http://www.knowledgetaiwan.org/ojs/index.php/ijbi/article/viewFile/342/96
39. Huang E, Cheng F (2012) Online security cues and e-payment continuance intention. Int J E-Entrepre Inno 3(1):42–58. https://doi.org/10.4018/jeei.2012010104
40. Atarodi S, Berardi AM, Toniolo AM (2018) The technology acceptance model since 1986: 30 years of development. Psychologie Du Travail et Des Organisations (Sept 2018). https://doi.org/10.1016/j.pto.2018.08.001
41. Basak SK, Govender DW, Govender I (2016) Examining the impact of privacy, Security, and trust on the TAM and TTF models for e-commerce consumers: a pilot study. In: 2016 14th Annual conference on privacy, security and trust, PST 2016, pp 19–26. https://doi.org/10.1109/PST.2016.7906922
42. Tsiakis T, Sthephanides G (2005) The concept of security and trust in electronic payments. Comput Secur 24(1):10–15. https://doi.org/10.1016/j.cose.2004.11.001

43. Mondego D, Gide E (2018) The effect of trust on mobile payment adoption: a comprehensive review of literature. Int J Arts Sci 11(1):375–389
44. Rouibah K, Lowry PB, Hwang Y (2018) The effects of perceived enjoyment and perceived risks on trust formation and intentions to use online payment systems: new perspectives from an Arab country. SSRN Electron J
45. Rose S, Hair N, Clark M (2011) Online customer experience: a review of the business-to-consumer online purchase context. Int J Manag Rev 13(1):24–39
46. Naseri MB, Elliott G (2011) Role of demographics, social connectedness and prior internet experience in adoption of online shopping: applications for direct marketing. J Target Meas Anal Mark 19(2):69–84
47. Hernández B, Jiménez J, Martín MJ (2010) Customer behavior in electronic commerce: the moderating effect of e-purchasing experience. J Bus Res 63(9–10):964–971
48. Rogers EM (2003) Diffusion of innovations, 5th edn. The Free Press, New York
49. Shaikh AA, Karjaluoto H (2015) Mobile banking adoption: a literature review. Telematics Inform 32(1):129–142
50. Sharma SK (2017) Integrating cognitive antecedents into TAM to explain mobile banking behavioral intention: a SEM-neural network approach. Inf Syst Front
51. Zhou T (2011) An empirical examination of initial trust in mobile banking. Internet Res 21 (5):527–540
52. Teoh WMY, Chong SC, Lin B, Chua JW (2013) Factors affecting consumers' perception of electronic payment: an empirical analysis. Internet Res 23(4):465–485
53. Esposito Vinzi V (2007) The contribution of PLS regression to PLS path modelling: formative measurement model and causality network in the structural model. Joint Statistical Meetings (JSM)
54. Wu JT, Leung K, Leung GM (2020) Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modeling study. Obstet Gynecol Surv 75(7):399–400
55. Sweeney JC, Danaher TS, McColl-Kennedy JR (2015) Customer effort in value cocreation activities: improving quality of life and behavioral intentions of health care customers. J Serv Res 18(3):318–335
56. Verleye K (2015) The co-creation experience from the customer perspective: its measurement and determinants. J Serv Manag 26(2):321–342
57. Pang C, Yang Q, Liu L, Yen DC, Michael Tarn J (2015) Exploring consumer perceived risk and trust for online payments: an empirical study in China's younger generation. Comput Hum Behav 50:9–24. https://doi.org/10.1016/j.chb.2015.03.058

Adam Optimized Deep Learning Model for Segmenting ROI Region in Medical Imaging

Mustafa Musa Jaber, Sura Khalil Abd, and Saif Mohammed Ali

Abstract Medical image processing played a crucial role in healthcare sectors because it produces visible images of inner body structure and internal tissues. With medical imaging, the organs' functions are analyzed, and anomalies are detected effectively. The imaging process utilizes enhancement, analysis, and recognition procedures that maximize disease detection accuracy. However, the gathered medical images have noise while performing the medical research that affects the ROI region segmentation process. The improper identification of the ROI region causes to reduce disease detection accuracy and create complexity. Therefore, this paper uses the Adam optimization technique with deep learning approaches for examining the medical images. The deep learning (DL) approaches utilizing the multiple layers, expert-tuned parameters, and learning function to deriving the affected ROT region. The DL method automatically extracts the medical imaging features that are optimized according to the RMS prop and momentum-based learning parameters. The optimized parameter minimizes the complexity while deriving the disease affected region. This process ensures high region segmentation accuracy with minimum time. The process has detected the affected region with 99.27% of accuracy.

Keywords Medical imaging · ROI region segmentation · Adam optimization technique · Deep learning model · Momentum learning parameter

M. M. Jaber (✉)
Information Technology Department, Dijlah University College, Baghdad, Iraq
e-mail: mustafa.musa@duc.edu.iq

M. M. Jaber
Al-Turath University College, Baghdad, Iraq

S. K. Abd
Computer Techniques Engineering, Dijlah University College, Baghdad, Iraq
e-mail: Sura.khalil@duc.edu.iq

S. M. Ali
Computer Science Department, Dijlah University College, Baghdad, Iraq
e-mail: saif.alameri@duc.edu.iq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_54

1 Introduction

Medical imaging [1, 2] is a technique that analyses and generates the internal body's visual representation for making the clinical analysis. The medical imaging process helps to identify the hidden internal structures to diagnose various diseases. This process represents the patient health conditions [3–5] in magnetic resonance imaging, X-ray radiography, tactile imaging, endoscopy, ultrasound etc. These radiological images examine the different pathological details in the earlier stage. The Adam optimization algorithm is an extension to stochastic gradient descent that has recently seen broader adoption for deep learning applications in computer vision and natural language processing [6]. The medical imaging has been performed by two types [7], such as invisible light and visible light procedures. The "invisible light" procedures are equivalent to the clinical or radiology images in which the practitioners involved in this analysis for recording the images. The "visible light" process, pictures or digital video has been perceived without having any equipment. In this work, "invisible lights" are passed to the patient body, and the radiology images are recorded with the help of radiologists in analyzing the various pathological details. According to the discussion, few radiology medical images are represented in Fig. 1.

Machine learning techniques [8] process the medical images such as radiation therapy, oncology, and radiology to make clinical decisions. In healthcare systems, artificial intelligence (AI) [9] and machine learning techniques are increased gradually and maximize the compound annual growth rate up to 40% in 2021. Moreover, the AI techniques are minimizing the computation complexity, reduce time consumption also less-expensive. Thus, AI technique and machine learning techniques are incorporated in the medical disciplines to analyze the radiology

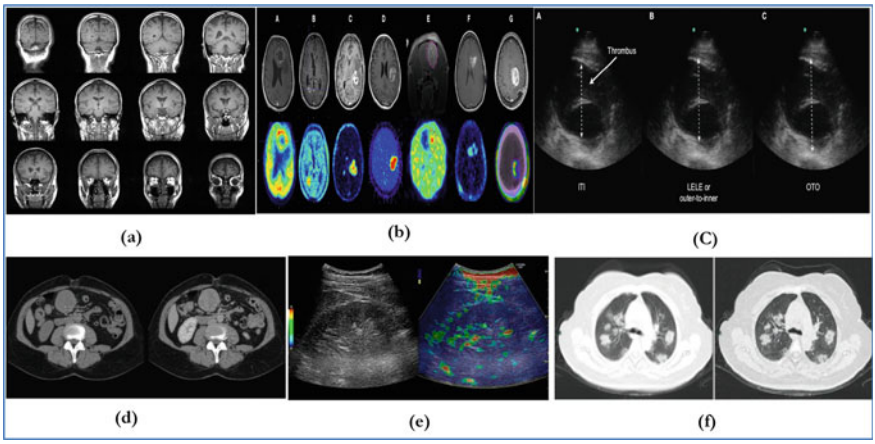


Fig. 1 Sample figures for medical imaging: **a** Magnetic resonance imaging **b** positron emission tomography **c** ultrasound **d** X-ray **e** elastography **f** computed tomography images

images [10–12]. The machine learning techniques has several algorithms such as supervised, statistical, unsupervised and semi-supervised algorithms are presented to examines the relationship between the data and recognize the hidden patterns effectively. This algorithm was interpreting the medical imaging process facing various processes while data acquisition also creating complexities when the number of data increases.

On the other hand, the traditional AI and machine learning techniques work according to various features, labor-intensive processes, and do not account for any learning or hidden patterns and time-consuming tasks. The traditional machine learning issues are overcome by applying deep learning (DL) concepts [13–16] to capture the hidden representations from the extracted features effectively. The maximum utilization of unstructured data, faster learning process, unwanted features elimination and minimum computation cost of the DL process improves the overall medical imaging analysis. The DL techniques consist of several methods [17, 18] such as Recurrent Neural Networks (RNN), Convolution Neural Networks (CNN), Long-Short Term Memory (LSTM) and Generative Adversarial Networks (GAN) which does not require the manual raw image processing. Among the techniques, Convolution Neural Networks (CNN) are utilized to successfully process the medical imaging because it recognizes the affected regions successfully. The CNN network minimizes features, preserving the most relevant features, eliminating the unwanted part and reducing the data overfitting. These characteristics are leads to improve the overall performance of the network in the medical imaging analysis.

Therefore, in this work, the CNN approach interprets and analyses the affected regions involved in medical imaging. In general, identifying the region of interest (ROI) [19] in the medical images is difficult because of hidden features. This causes to minimize the disease detection accuracy and time-consuming task. Hence, this algorithm, Adam optimization techniques are incorporated with deep learning technique (DL) to improve the overall ROI region identification accuracy. The optimization algorithm works to replace the stochastic gradient descent approach [20], enhancing the overall DL training process. The algorithm then has RMS prop and momentum-based learning parameters that effectively handle the sparse gradient on the noisy problem in medical images and easy to configure with DL networks. Due to these reasons, in this work, Adam optimization algorithm is working with the deep learning (DL) process for analyzing the medical images. Thus, the created system minimizing the computation complexity, time-consumption and improve the overall recognition accuracy effectively. The discussed system evaluated using MATLAB implementation tool with respective performance metrics. The manuscript is then organized as follows: Sect. 2 reviews the different techniques, research thoughts, and ideas regarding medical imaging analysis. Section 3 discuss the working process of optimized DL based medical imaging analysis and effectiveness is evaluated in Sect. 4. Summarization of this work is discussed in Sect. 5.

2 Related Works

This section discusses the various researcher opinions, techniques and ideas regarding the medical imaging analysis process. This research [21] creating the medical imaging analyzing and synthesis system for segmenting the affected region by applying the adversarial confidence learning. This process uses the fully convolutional adversarial network (FCAN) to examine the region's images and voxel-wise. From the examined details, structural information's are obtained to identify the irregular circulation of medical data. According to this analysis, the affected regions are extracted, improving the entire system performance in quantitative and visual perception.

Reference [22] analyzing the bone-cancer and destruction patterns for X-ray image using support vector machine (SVM). The long bone X-ray images are continuously investigated according to the digital geometry property. The SVM approach gets the cancer-related details passed to the decision tree classifier to identify the destruction patterns from the obtained details.

Reference [23] utilizing the group normalized convolution network (GnCNNr) with the regulation approach to identify the musculoskeletal abnormalities. This work used the MURA dataset, which consists of forearm, shoulder, elbow, finger, wrist and hand-related details. The dataset has around 14,683 images which are examined using deep learning model (DL). The DL model uses the weight standardization, group normalization and cyclic learning rate for analyzing the changes in the medical images. The introduced method ensures 85% accuracy and 0.698 Cohen Kappa statistic values higher than other deep learning techniques.

Reference [24] analyzing the MRI brain images using different deep learning techniques (DL). The DL techniques are utilized to investigating the medical images according to the contexts, and the relevant clinical information's are obtained to making the clinical decisions. This research work uses the almost 427 publications from Science direct, Scopus database and IEEEXplore Library of Engineering database. From the various applications, DL is utilized to examine the medical images for identifying further clinical analysis.

Reference [25] reviewing the deep learning model to examining the medical images. This work examines the structure of convolution networks, transfer learning strategies, knowledge of transfer learning and systematic view of medical images related transfer learnings are discussed. From the discussions, deep learning and effective transfer learning is selected to examine the various medical images.

Reference [26] Applying deep learning and feature enhancement technique to segmenting the affected region from medical images. This work aims to reduce the difficulties in edge and texture related information selection process. For this purpose, the Gabor network-based deep learning (AGnet) approach is utilized to analyze the edge relevant details. Initially, the images are analyzed using convolution filters to identify the approximate locations. Then the Gabor filters are applied to extract the salient edge relevant features. From the derived information, disease-affected regions are segmented successfully.

Reference [27] analyzing the different meta-heuristic optimization algorithms for segmenting the medical images. This work was investing the other soft computing techniques such as memetic algorithm, ant colony optimization, particle swarm, evolutionary computation, tabu search, fuzzy systems and stimulated annealing methods to identify the affected region. The working process and the segmenting procedures are discussed for improving the overall medical region segmentation process.

Reference [28] Investigating the multimodal medical imagining by applying the deep learning model to perform the segmentation. The medical images are analyzed according to the DL network structure called cross-modality fusion. This process examines the multimodal images of contour related details. From the contour information, changes involved in the pixels are identified and grouped effectively.

Reference [29] Segmenting the human thyroid medical images by applying the dual-tree complex wavelet transform with convolution neural networks (DWT-CNN). The introduced approach resolves the limitations involved in the textural information identification and insufficient robustness to the noise. Here, the dual-tree complex wavelet approach is integrated with the convolution networks pooling layer to examine the image pixels. This process extracts the texture features in the spatial directions by decomposing the images into various sub-bands. Then the high-frequency sub-bands are removed from the analysis for eliminating the noise details. This process performed continuously to segment the thyroid region from the human thyroid optical coherence tomography images.

According to the various researcher's opinions, the medical images are examined by applying the different deep learning model. The DL model investigated successfully every pixel in the image according to the transfer learning and activation function. Although the medical images are captured according to the user intervention, therefore, noise may have occurred. The noise involved in the medical images causes to identify the wrong region of interest (ROI). Thus, the ROI analysis process should be improved by reducing the computation complexity. Hence, the optimized deep learning model is introduced to improve the overall medical image analysis accuracy.

3 Medical Image Region Analysis Using Optimized Deep Learning Model

This section is discussing the working process of optimized deep learning model based medical image region analysis process. This system uses three-steps such as preprocessing, region segmentation and post-processing to analyzing the medical images for making the clinical decisions. Then the detailed working process of the medical image region analysis system illustrated in Fig. 2.

Figure 2 illustrated that the medical imaging region analysis working process structure. The optimized deep learning model (ODL) examines the disease affected

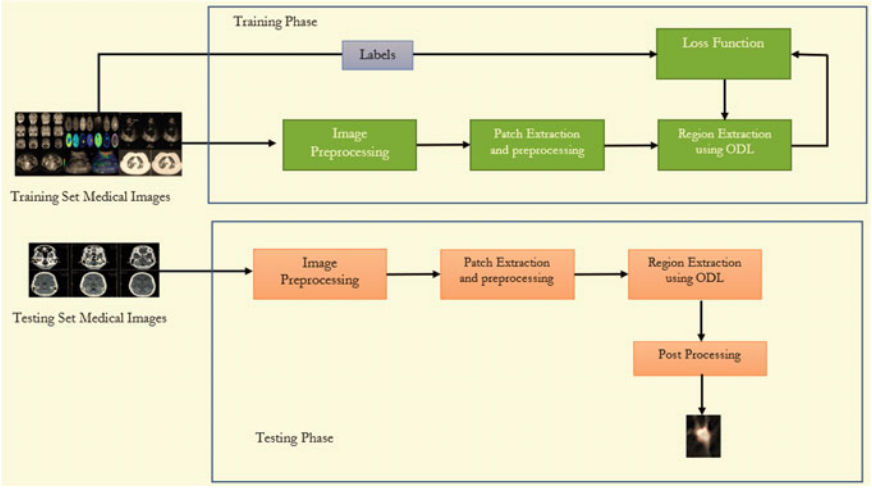


Fig. 2 Medical image region analysis structure

region in two phases, such as training and testing. In the training phase, medical imaging with labels is given to the system that processes the images using image preprocessing, patch extraction, and segmentation. This training process examines each pixel in the image and respective labels according to the pixel characteristics. Then the testing phase, the test images are processed by the same steps and the segmentation regions are identified according to the given labels. From the detected region post-processing is performed to get the disease-relevant details.

The medical imaging details and respective collected details are illustrated in Table 1.

From Table 1, it clearly shows that the dataset has several medical imaging information. These information's are used to process by ODL approach to predict the disease affected region. The dataset consists of almost MRI and CT images; these images are processed according to the preprocessing, patch extraction and segmentation steps to derive the disease affected region. Then the detailed working process is explained in the following section.

3.1 Dataset Description

In this work, Medical Segmentation Decathlon dataset images [30] are utilized to examine the introduced ODL method excellence. The dataset consists of various medical image information such as brain tumour, heart, liver, hippocampus: prostate, lung, pancreas, Hepatic Vessel, Spleen and colon.

Table 1 Dataset description

Medical imaging	Parameters	Information
Liver tumour	Target	Liver and tumour
	Modality	Portal venous phase CT
	Size	201 3D volumes (70 testing and 131 training image)
	Challenge	Label: unbalance with a large and small tumour
Brain tumor	Target	Gliomas segmentation of necrotic and oedema
	Modality	Multimodal multisite MRI image
	Size	750 4D volumes (266 testing and 484 training images)
	Challenge	Heterogeneously and complex located targets
Hippocampus	Target	Body and Hippocampus
	Modality	Mono-modal MRI
	Size	394 3D volumes (131 testing and 263 training images)
	Challenges	Segmenting into small structures of two neighboring details
Lung tumor	Target	Lung tumour
	Modality	CT
	Size	96 3D volumes (32 testing and 64 training images)
	Challenges	Segmenting cancer from a large image
Prostate	Target	Peripheral zone and central prostate gland
	Modality	Multimodal MR images
	Size	48 4Dvolumes (16 testing and 32 training images)
	Challenges	Segmenting two adjoint regions
Cardiac	Target	Left Atrium
	Modality	Mono-modal MRI images
	Size	30 3D volumes (10 testing and 20 training images)
	Challenges	Small training set with large variability
Pancreas tumor	Target	Liver and tumour
	Modality	Portal venous CT images
	Size	420 3D volumes (139 testing and 282 training images)
	Challenges	Label with a large, medium and small structure
Colon cancer	Target	Colon cancer
	Modality	CT images
	Size	190 3D volumes (64 Testing and 126 training images)
	Challenges	Heterogenous appearance
Hepatic vessels	Target	Hepatic vessels and tumour
	Modality	CT images
	Size	443 3D volumes (140 Testing and 303 Training images)
	Challenges	Tubular small structures to the heterogeneous tumour
Spleen	Target	Spleen
	Modality	CT
	Size	61 3D volume (20 Testing and 41 Training images)
	Challenges	Large ranging foreground size

3.1.1 Medical Image Preprocessing

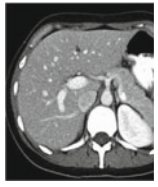
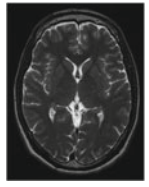
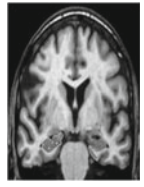
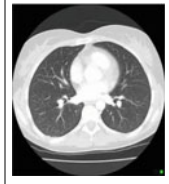
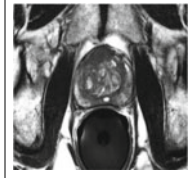
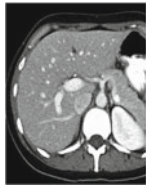
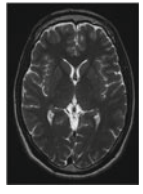
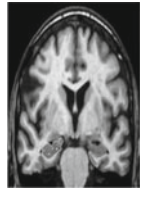
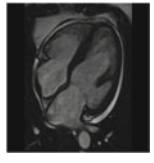
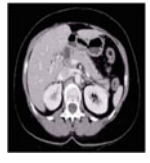
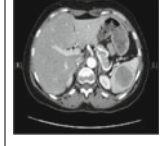
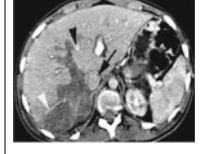
The first step of this process is to eliminate the unwanted noise from the medical images. The medical images are affected by different noises such as salt pepper, gaussian, speckle, poison and blurred noise. This step attempts to eliminate the unwanted noise from the medical image and improve the image quality for further clinical analysis. The noise removal process removed the image's noise by restoring the pixel values—almost every medical image affected by Gaussian noise (GN). The GN is nothing but the gaussian distribution whose PDF (probability distribution function) is equal to the statistical noise. The GS noise is eliminated by performing the smoothening image process that minimizes the noisy pixel intensity value. This GN noise causes the blurring edges in the high-quality images. Therefore, the GN noise should be removed from the medical images for improving the quality of the clinical analysis. In this work, the median filter is utilized to eliminate the noise from the medical image. It is one of the effective techniques and having good de-noising power with mathematical accuracy. Each pixel in the medical images is needed to examine with the intensity of neighborhood pixel value. Suppose the pixel has the different intensity values which are replaced by computing the median value. During the computation process, the fixed window size is utilized to estimate the median pixel value. This is the simple and most comfortable task for eliminating the noise also increase the quality of the image. According to the explanation, the sample medical and noise removed medical images are illustrated in Table 2.

After eliminating the image's noise, mean, standard deviation, an intensity value and variance values are computed for every pixel in the image. According to these values, the image patches are identified based on the intensity value because here, the remaining statistical measure is normalized to zero. Once the image patches are identified then again, median filters are applied to examining the noise. Then the noise removed images are fed into the segmentation procedure to identify the disease affected region.

3.1.2 Medical Imaging Segmentation

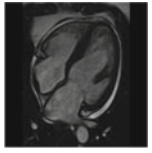
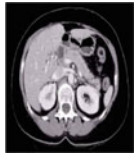
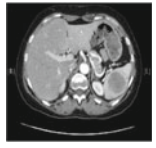
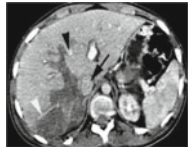
This section explains the medical image segmentation process by applying an optimized deep learning model (ODL). Image segmentation is nothing but the procedure of dividing the medical images into different regions or segments. The segmentation process changes the representation of the original images into meaningful sub-images. This process simplifies the entire process also make it easier to analyses. The medical image segmentation process helps to determine the regions, boundaries, curves and lines etc. From the identified information, abnormal regions are predicted effectively by using the pixel characteristics. In this work, Adam optimized the convolution neural network to segment the disease affected the region from the various medical images. The convolution neural network is one of the effective segmentation approaches which is done by using the different

Table 2 Sample and noise removed medical images

Images	Liver tumor	Brain	Hippocampus	Lung	Prostate
Sample images					
Noise removed images					
Images	Cardiac	Pancreas	Colon	Spleen	Hepatic vessels
Sample images					

(continued)

Table 2 (continued)

Images	Cardiac	Pancreas	Colon	Spleen	Hepatic vessels
Noise removed images					

convolution layers. The networks have a feature map, in which kernels are utilized to mapping each pixel with the respective feature map. Each kernel has weight value used to connect the feature map with the previous layers—the network weight value used to train the image and backpropagation process to reduce the error rate. The convolution layers, feature map and pooling layers are utilized to minimize the feature overfitting also train the image patterns quickly. During the image segmentation process, non-linear activation functions are being used to compute the pixel's output.

3.1.3 Architecture of Convolution Neural Networks

The convolution neural network architecture is illustrated in Fig. 3. Generally, the medical images have the enormous variability in their intra-tumoral structures which create the ambiguity during the edge's analysis. This segmentation problem causes computation complexity; therefore, the convolution networks are designed to resolve the above-discussed issue.

Figure 3 shows the convolution network structure with the input layer, feature mapping, convolution layer, max pooling, and output layer. The network consumes the medical images as input, and the segmentation is performed by sensing the exact disease affected location. The pooling process reduces irrelevant features and minimizes the data overfitting. This is achieved by applying 3 * 3 receptive fields of overlapping pooling process and 2 * 2 strides for locating region. Before using the convolution, operator features maps are padded in the convolution layer, maintaining the image's dimension. The ReLu activation function is utilized to compute the output value, and the final layer uses the SoftMax as the activation function. Moreover, the network performance of the system is improved by applying the Adam optimization technique. The convolution network utilized parameters are illustrated in Table 3.

Table 3 shows that the input has the three dimensions in which the first dimension is represented as the number channels and the remaining two dimensions denoted as the size of patches or feature maps. After defining the CNN network

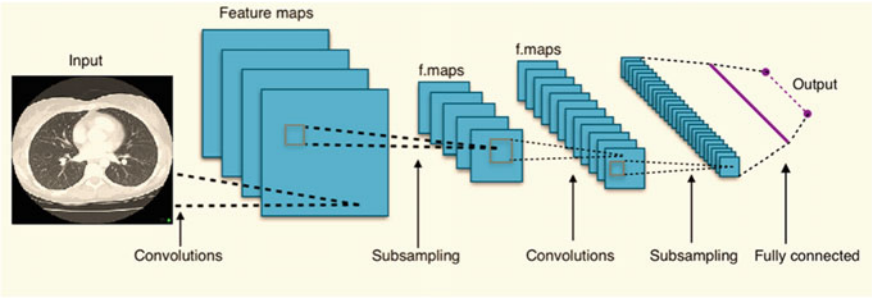


Fig. 3 Convolution neural network structure

Table 3 Convolution neural network architecture

Layer	Type	Input	Filter size	# filters	FC unit	Stride
Layer 1	Conv…	4 * 33 * 33	3 * 3	64	–	1*1
Layer 2	Conv…	64 * 33 * 33	3 * 3	64	–	1 * 1
Layer 3	Conv…	64 * 33 * 33	3 * 3	64	–	1 * 1
Layer 4	Max pool	64 * 33 * 33	3 * 3	–	–	2 * 2
Layer 5	Conv…	64 * 16 * 16	3 * 3	128	–	1 * 1
Layer 6	Conv…	128 * 16 * 16	3 * 3	128	–	1 * 1
Layer 7	Conv…	128 * 16 * 16	3 * 3	128	–	1 * 1
Layer 8	Max pool	128 * 16 * 16	3 * 3	–	–	2 * 2
Layer 9	FC	6272	–	–	256	–
Layer 10	FC	256	–	–	256	–
Layer 11	FC	256	–	–	5	–

*Conv.—convolution layer, Max pool—Max pooling layer and FC—fully connected layer

architecture, the initialization should be performed to ensures the convergence. Here, Xavier initialization is achieved, which means, the gradient and activation functions are controlled else the gradients are exploded or vanish. This segmentation process uses the rectified linear unit (ReLU) as an activation function that is defined in Eq. (1)

$$f(x) = \max(0, x) \tag{1}$$

The ReLU function speeds up the medical image training process in which 0 value indicates that the network requires the consequential adjustment of weight values. The pooling should then be performed in which the feature map is applied to combine the spatial features and reduce the feature redundancy. The pooling process reduces the computational load; mostly, the max-pooling function reduces the feature size. Then the feature overfitting should be performed via the regularization process. Here, drop out function is applied in the FC layer, which is done by according to the pixel probability value. After reducing the feature size, data augmentation should be performed that increasing the training set and minimize the overfitting issues. The augmentation process has rotating operations; here 90° is utilized as the rotation angle. Based on the angle, medical image data has been analyzed, and segmentation is performed. As discussed earlier, the network substantial network weight value is updated according to the estimated loss function. Loss value is minimized during data training. Here, cross-entropy value is used to compute the loss function defined in Eq. (2).

$$H = -\sum_{j \in voxel} \sum_{k \in classes} c_{j,k} \log\left(\hat{c}_{j,k}\right) \tag{2}$$

Probabilistic prediction of the affected region is denoted as $\widehat{c}$ and the target value is c. This computed loss function should be minimized while training medical images. Here, Adam optimized algorithm is utilized to enhance the overall performance of the convolution neural networks. The advanced version of the stochastic gradient descent algorithm is used to minimize the deviation of estimated and computed output value. It works a mixture of root mean square (RMS) prop and momentum learning process. This learning parameter accumulates the gradient of previous steps to reach the global minima value. Along with the stochastic approach, the momentum parameter is utilized to improve the feature training process. Then the learning parameter terms are defined via Eq. (3).

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta); \theta = \theta - v_t \tag{3}$$

From the variant values defined in Eq. (3), the Adam optimization algorithm related training parameters are defined in Eqs. (4 and 5)

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{4}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{5}$$

In Eqs. (4 and 5), moving average gradients are denoted as m and v. β is used as Adam optimization parameter that has value $\beta_1 = 0.9$ and $\beta_2 - 0.999$; g is the gradient in mini-batch and momentum term is $\gamma = 0.9$.

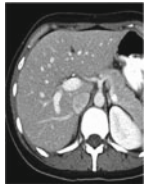
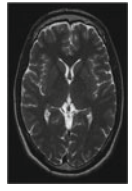
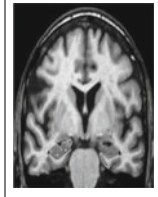
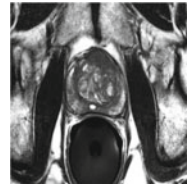
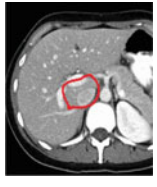
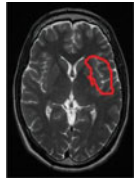
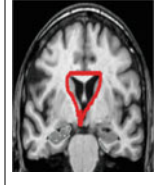
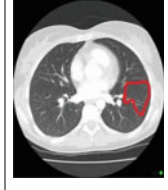
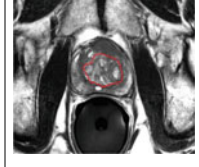
Based on the above learning process, each pixel is examined in the medical image, and the small clusters are formed according to the pixel information. This process is repeated in the post-processing to group the affected region successfully. According to the above discussion, the formed clusters in medical images are shown in Table 4.

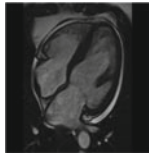
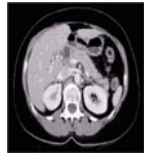
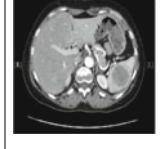
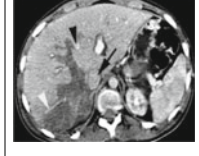
The segmented regions are further analyzed clinical researchers in taking the clinical decisions. According to the above discussion, the convolution layer examines the image pixel according to the optimized learning parameters to effectively derive the affected region.

4 Results and Discussions

This section evaluates the excellence of the Optimized Deep Learning (ODL) model-based medical image segmentation process. This process utilizes the Medical Segmentation Decathlon dataset images for examining the proficiency of the system. The detailed description of the dataset is illustrated in Table 1. The collected medical images are implemented using MATLAB tool and the different performance metrics such as pixel accuracy (PA), Mean Pixel Accuracy (MPA), Interaction over Union (IoU), Dice Coefficient and segmentation accuracy. The dataset consists of around ten disease images which are processed by training and

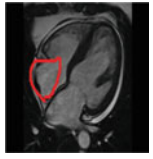
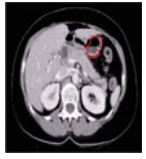
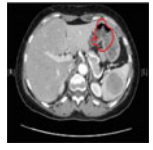
Table 4 Medical image segmented region analysis

Images	Liver tumor	Brain	Hippocampus	Lung	Prostate
Noise removed images					
Segmented images					

Images	Cardiac	Pancreas	Colon	Spleen	Hepatic vessels
Noise removed images					

(continued)

Table 4 (continued)

Images	Cardiac	Pancreas	Colon	Spleen	Hepatic vessels
Segmented images					

testing phase details. The noise presented in the images is removed by computing the pixel intensity value and the respective median value. Then different convolution layers are applied to segment the affected medical region. For each image, the training and testing images are defined in Table 1, the effective training process minimizes the loss function and improves the overall segmentation accuracy.

4.1 Performance Metrics and Analysis

Pixel Accuracy (PA)

PA is nothing but the ratio of the pixels which are classified from the whole number of images. The PA is computed using Eq. (6)

$$PA = \frac{\sum_{i=0}^{K} p_{ij}}{\sum_{i=0}^{K} \sum_{j=0}^{K} p_{ij}} \tag{6}$$

In Eq. (6), pixels belonging to the i class is defined as p_{ij} and predicted class is j.

From the computation of PA in Eq. (6), the medical image's pixels are examined effectively. The ODL technique obtained results are compared with the existing researchers works such as a fully convolutional adversarial network (FCAN) [21], group normalized convolution network (GnCNNr) [23], Gabor network-based deep learning (AGnet) [26], dual-tree complex wavelet transforms with convolution neural networks (DWT-CNN) [29]. These methods are applied to the Medical Segmentation Decathlon dataset images, and the obtained PA values are illustrated in Table 5.

Table 5 Pixel accuracy (PA) (%)

S. No	Images	FCAN	GnCNNr	AGnet	DWT-CNN	ODL
1	Liver tumor	94.78	95.89	96.29	97.8	98.92
2	Brain	94.79	95.62	96.72	97.52	98.10
3	Hippocampus	95.02	95.91	97.02	97.91	99.29
4	Lung	94.89	96.02	96.59	97.64	98.98
5	Prostate cancer	94.91	95.89	96.921	97.28	99.34
6	Cardiac	95.29	96.28	96.28	97.45	99.53
7	Pancreas	95.72	96.62	97.29	97.28	99.37
8	Colon	94.98	95.92	97.39	97.673	99.62
9	Spleen	94.87	95.28	97.32	97.68	99.36
10	Hepatic vessels	95.02	96.28	97.03	98.29	99.189

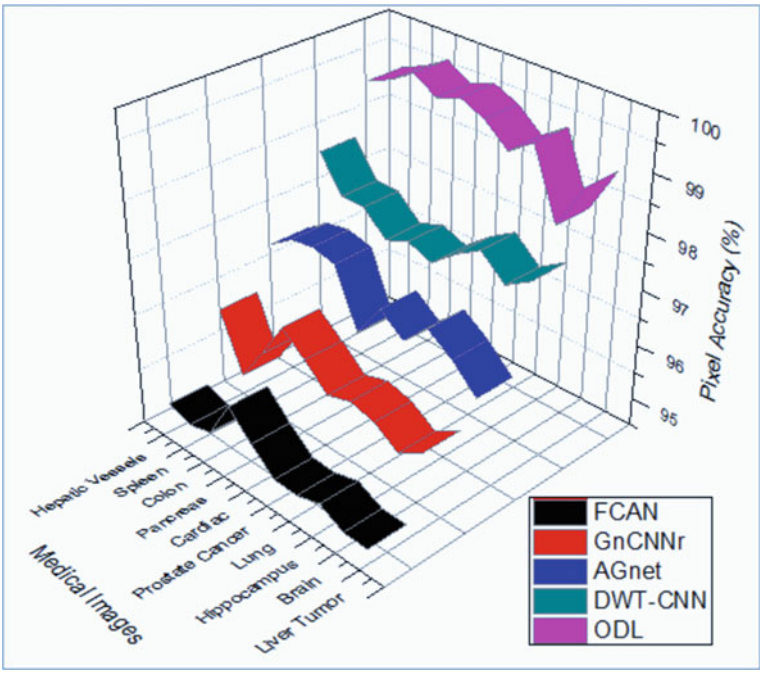


Fig. 4 Pixel accuracy

Table 5 illustrated that the Pixel Accuracy value of different medical image segmentation methods. From the analysis, the introduced ODL approach attains the high image segmentation accuracy. The image pixels are examined by mapping the features with the respective features using convolution operators and strides.

The filters and max-pooling process minimize the data over fittings, which leads to improving the pixel analysis rate. According to the results, the graphical evaluation of pixel accuracy is shown in Fig. 4.

Figure 4 illustrated that the pixel accuracy value of various medical image region analysis. The introduced ODL approach attains the high PA value (99.16%) compared to the other methods such as FCAN (95.02%), GnCNNr (95.97%), AGnet (96.88%) and DWT-CNN (97.65%). Therefore, the ODL approach examines each pixel presented in the medical image effectively. After reviewing the pixels, the disease affected the right pixels must be selected from the entire pixel. This has to be evaluated using the mean pixel accuracy (MPA) value. The MPA value computed using Eq. (7).

$$MPA = \frac{1}{K+1}\sum_{i=0}^{K}\frac{p_{ij}}{\sum_{j=0}^{K} p_{ij}} \tag{7}$$

Table 6 Mean pixel accuracy (MPA) (%)

S. No	Images	FCAN	GnCNNr	AGnet	DWT-CNN	ODL
1	Liver tumor	95.29	95.88	97.29	98.8	99.43
2	Brain	95.113	96.24	97.83	98.24	99.56
3	Hippocampus	95.27	96.38	97.62	98.65	99.36
4	Lung	95.87	96.83	97.91	98.48	99.52
5	Prostate cancer	95.29	96.92	97.32	98.62	99.47
6	Cardiac	95.92	96.83	97.69	98.38	99.62
7	Pancreas	95.03	96.91	97.92	98.42	99.65
8	Colon	95.72	96.53	97.58	98.47	99.38
9	Spleen	95.27	96.28	97.82	98.56	99.15
10	Hepatic vessels	95.78	95.98	97.85	98.67	99.23

From the computation of MPA in Eq. (6), the medical image's pixels are examined and identified the right pixels. The ODL technique obtained MPA results are compared with the existing [21, 23, 26, 29]. These methods are applied to the Medical Segmentation Decathlon dataset images, and the obtained MPA values are illustrated in Table 6.

Table 6 illustrated that the Mean Pixel Accuracy (MPA) value of different medical image segmentation methods. From the analysis, the introduced ODL approach attains the high image segmentation accuracy. The pixels are examined, and changes in the pixels are determined according to the max-pooling and data augmentation process. The right pixels are selected, and the deviation of the pixels is reduced by computing the loss function. The momentum and gradient learning parameters updating process causes to minimize the variation of the segmentation accuracy. Then the respective MPA graphical analysis is shown in Fig. 5.

Figure 4 illustrated that the pixel accuracy value of various medical image region analysis. The introduced ODL approach attains the high MPA value (99.43%) compared to the other methods such as FCAN (95.45%), GnCNNr (96.47%), AGnet (97.68%) and DWT-CNN (98.52%). Therefore, the ODL approach examines each pixel presented in the medical image effectively. Further, the segmentation process's excellence is evaluated using the Intersection over Union (IoU) metric. This metric used to identify how effectively, the introduced ODL approach detects the medical image and the segmented region. The IoU metric is computed using Eq. (8).

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{8}$$

The IoU has also named the Jaccard Index computed from the identified segmented region and the ground truth map. A, and B are the forecast and ground truth segmentation map details. According to the Eq. (8), the computed IoU value related graphical analysis shown in Fig. 6.

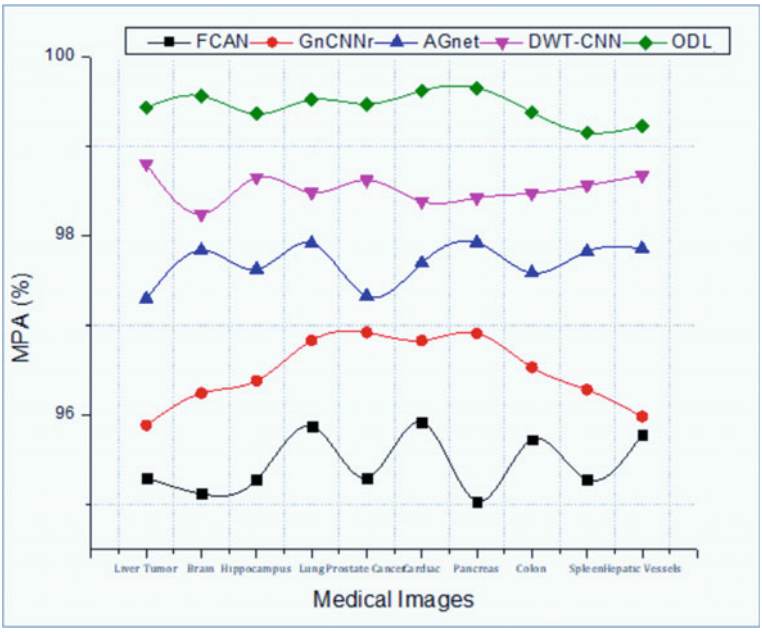


Fig. 5 Mean pixel accuracy

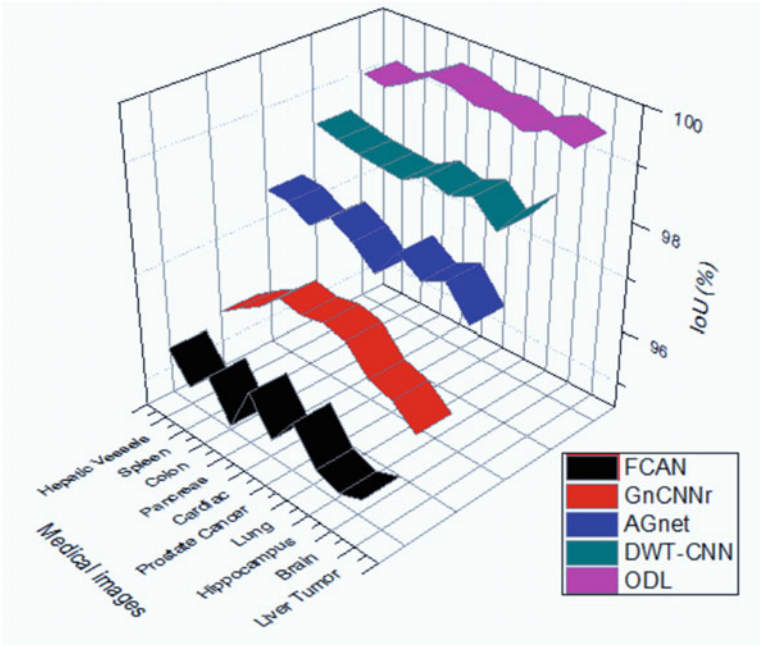


Fig. 6 IoU analysis

Figure 6 illustrated that the obtained IoU value of different medical image analysis process. The introduced system removes the noise from the initial stage itself, leading to the effective treatment of the disease region. Then the affected area related features are computed with minimum loss function $H = -\sum_{j \in voxel} \sum_{k \in classes} c_{j,k} \log\left(\hat{c}_{j,k}\right)$ and the process is further optimized using the momentum parameter $v_t = \gamma v_{t-1} + \eta \nabla_\theta J(\theta); \theta = \theta - v_t$. This process helps to reduce the error rate in the training stage itself. The successful training process causes to segment the testing images successfully. Further, the extracted region similarity to the training sample and the disease affected region is evaluated using the Dice coefficient (DC). The DC value is estimated using Eq. (9).

$$DC = \frac{2|A \cap B|}{|A| + |B|} \tag{9}$$

According to Eq. (9), the computed output is shown in Fig. 7. The DC values are examined to the different medical image region segmentation process, which is compared with various methods such as [21, 23, 26, 29].

Figure 7 illustrated that the obtained DC value of different medical image analysis process. The introduced system removes the noise using the median filter that smoothens the image, leading to successfully examining the region-related edges. Then the affected area related features are computed with a minimum loss

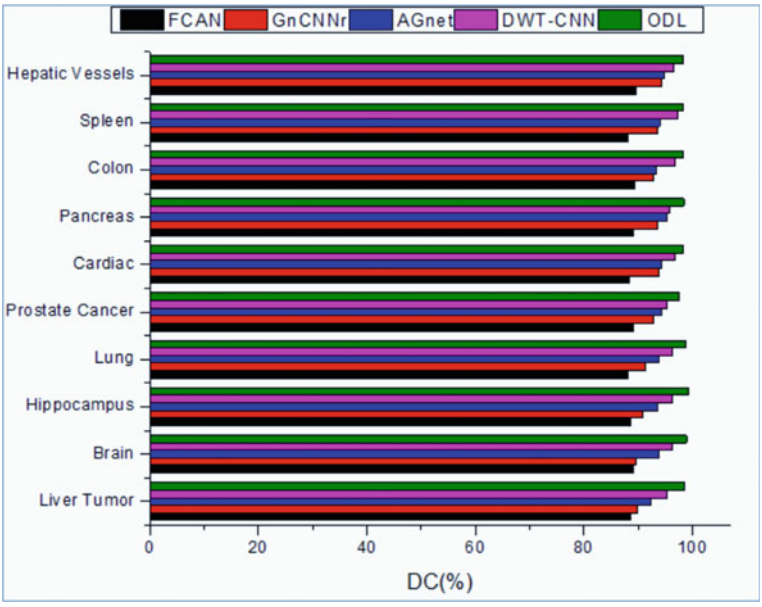


Fig. 7 DC analysis

Table 7 Segmentation Accuracy

S. No	Methods	Training	Testing	Overall accuracy
1	Fully Convolutional Adversarial Network (FCAN)	96.279	97.289	96.784
2	Group Normalized Convolution Network (GnCNNr)	97.29	97.83	97.56
3	Gabor network-based deep learning (AGnet)	97.82	98.29	98.055
4	Dual-Tree Complex Wavelet Transforms With Convolution Neural Networks (DWT-CNN)	98.34	98.48	98.41
5	Optimized Deep learning model (ODL)	99.234	99.31	99.272

function $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$ and $v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$ and the process is further optimized using the momentum parameter. The entire PA, MPA, IoU and DC metrics show that the method improves the overall segmentation accuracy. Then the comprehensive training and testing process of introducing method accuracy illustrated in Table 7.

From the Table 7 clearly shows that the ODL technique obtained high segmentation accuracy (99.27%) compared to existing researchers works fully convolutional adversarial network (FCAN) [96.78%], group normalized convolution network (GnCNNr) [97.576%], Gabor network-based deep learning (AGnet) [98.05%], dual-tree complex wavelet transforms with convolution neural networks (DWT-CNN) [98.41]. These methods effectively attain the high results on both training, testing and overall accuracy. Thus the introduced system successfully recognize the disease affected region using different convolution layers and respective learning parameters.

5 Conclusion

Thus the paper analyzing the optimized deep learning model (ODL) method based medical imaging analysis process. Initially, the medical images are collected from the Medical Segmentation Decathlon dataset that consists of different medical images. The noise involved in the images is eliminated by computing the pixel intensity medium value. This process reduces the noise and smoothens the image by retaining the edge relevant features. The affected region corresponding pixels are then identified using the feature map, max pooling, and convolution layer process. During the analysis, the loss function is reduced by updating the network weight value according to the Adam optimization learning parameters such as momentum variance. This process effectively detects the affected region with 99.27% of accuracy. In future, the excellence of the system is improved by integrating the meta-heuristic optimization technique with deep learning model.

References

1. Giger ML (2018) Machine learning in medical imaging. J Am Coll Radiol 15(3):512–520
2. Zhou Z, Siddiquee MM, Tajbakhsh N, Liang J (2018) Unet++: a nested u-net architecture for medical image segmentation. In: Deep learning in medical image analysis and multimodal learning for clinical decision support, pp 3–11. Springer, Cham
3. Iaquinta P, Iusi M, Caroprese L, Turano S, Palazzo S, Dattola F, Pellegrino I, Veltri P, Zumpano E (2017) eIMES 3D: an innovative medical images analysis tool to support diagnostic and surgical intervention. Procedia Comput Sci 110:459–464
4. Na Y, Zhao L, Yang Y, Ren M (2017) Guided filter-based images fusion algorithm for CT and MRI medical images. IET Image Proc 12(1):138–148
5. Meng L, Guo X, Li H (2019)MRI/CT fusion based on latent low rank representation and gradient transfer. Biomed Sig Process Control 53:101536
6. Jais IKM, Ismail AR, Nisa SQ (2019) Adam optimization algorithm for wide and deep neural network. Knowl Eng Data Sci
7. Mairhofer S, Zappala S, Tracy SR, Sturrock C, Bennett M, Mooney SJ, Pridmore T (2012) RooTrak: automated recovery of three-dimensional plant root architecture in soil from X-ray microcomputed tomography images using visual tracking. Plant Physiol 158(2):561–569
8. Ker J, Wang L, Rao J, Lim T (2017) Deep learning applications in medical image analysis. IEEE Access 6:9375–9389
9. Savadjiev P, Chong J, Dohan A, Vakalopoulou M, Reinhold C, Paragios N, Gallix B (2019) Demystification of AI-driven medical image interpretation: past, present and future. Eur Radiol 29(3):1616–1624
10. Karimi D, Dou H, Warfield SK, Gholipour A (2020) Deep learning with noisy labels: exploring techniques and remedies in medical image analysis. Med Image Anal 65:101759
11. Litjens G, Kooi T, Bejnordi BE, Setio AA, Ciompi F, Ghafoorian M, Van Der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. Medical image analysis 42:60–88
12. Kazeminia S, Baur C, Kuijper A, van Ginneken B, Navab N, Albarqouni S, Mukhopadhyay A (2020) GANs for medical image analysis. Artif Intell Med 109:101938
13. Verma OP, Roy S, Pandey SC, Mittal M (eds) (2019) Advancement of machine intelligence in interactive medical image analysis. Springer Nature
14. Mittal M, Arora M, Pandey T, Goyal LM (2020) Image segmentation using deep learning techniques in medical images. In: Advancement of machine intelligence in interactive medical image analysis, pp 41–63. Springer, Singapore
15. Vasconcelos FF, Sarmento RM, Reboucas Filho PP, de Albuquerque VH (2020) Artificial intelligence techniques empowered edge-cloud architecture for brain CT image analysis. Eng Appl Artif Intell 91:103585
16. Smailagic A, Costa P, Gaudio A, Khandelwal K, Mirshekari M, Fagert J, Walawalkar D et al (2020) O-MedAL: Online active deep learning for medical image analysis. Wiley Interdiscip Rev: Data Min Knowl Discov 10(4):e1353
17. Zhang Z, Cui P, Zhu W (2020) Deep learning on graphs: a survey. IEEE Trans Knowl Data Eng
18. Ahmad J, Farman H, Jan Z (2019) Deep learning methods and applications. In: Deep learning: convergence to big data analytics, pp 31–42. Springer, Singapore
19. Pandey D, Yin X, Wang H, Su MY, Chen JH, Wu J, Zhang Y (2018) Automatic and fast segmentation of breast region-of-interest (ROI) and density in MRIs. Heliyon 4(12):e01042
20. Yang J, Yang G (2018) Modified convolutional neural network based on dropout and the stochastic gradient descent optimizer. Algorithms 11(3):28
21. Nie D, Shen D (2020) Adversarial confidence learning for medical image segmentation and synthesis. Int J Comput Vis 128:2494–2513. https://doi.org/10.1007/s11263-020-01321-2

22. Bandyopadhyay O, Biswas A, Bhattacharya BB (2019) Bone-cancer assessment and destruction pattern analysis in long-bone x-ray image. J Digit Imaging 32:300–313. https://doi.org/10.1007/s10278-018-0145-0
23. Goyal M, Malik R, Kumar D et al (2020) Musculoskeletal abnormality detection in medical imaging using GnCNNr (Group normalized convolutional neural networks with regularization). SN Comput Sci 1:319. https://doi.org/10.1007/s42979-020-00340-7
24. Al-Galal SAY, Alshaikhli IFT, Abdulrazzaq M (2021) MRI brain tumor medical images analysis using deep learning techniques: a systematic review. Health Technol. https://doi.org/10.1007/s12553-020-00514-6
25. Wang J, Zhu H, Wang SH et al (2020) A review of deep learning on medical image analysis. Mobile Netw Appl. https://doi.org/10.1007/s11036-020-01672-7
26. Huang S, Huang M, Zhang Y, Chen J, Bhatti U (2020) Medical image segmentation using deep learning with feature enhancement. IET Image Process 14(14):3324–3332
27. Dey N, Ashour AS (2018) Meta-heuristic algorithms in medical image segmentation: a review. In: Advancements in applied metaheuristic computing, pp 185–203. IGI Global
28. Guo Z, Li X, Huang H, Guo N, Li Q (2019) Deep learning-based image segmentation on multimodal medical imaging. IEEE Trans Rad Plasma Med Sci 3(2):162–169
29. Lu H, Wang H, Zhang Q, Won D, Yoon SW (2018)A dual-tree complex wavelet transform based convolutional neural network for human thyroid medical image segmentation. In: 2018 IEEE international conference on healthcare informatics (ICHI), pp 191–198. IEEE
30. http://medicaldecathlon.com/

Analyzing the Digital Pathology Images Using Optimized Generative Adversarial Deep Learning Networks

Sura Khalil Abd, Mustafa Musa Jaber, and Saif Mohammed Ali

Abstract In recent years, digital pathology requires continuous detection and progress to reduce risk factors in medical applications. The deep learning (DL) model's improvement reduces digital pathology's complexity and minimizes laborious modelling. The DL modelling was investigating the whole slide images for identifying the lymph node metastases and tumor regions. Even though the DL method requires advanced optimization methodologies to improve the pathology detection performance, the traditional DL model faces difficulties in quantifying the mitoses and inflammation specific features. The irrelevant features and segmented regions reduce the performance of digital pathology detection. Therefore, Prewitt kernel with a butterfly-optimized, generative adversarial deep learning network (PK-BGAN) performs the digital pathology process in this work. The Prewitt kernel process examines each pixel in the medical images that help to extract the changes involved in the image. Significant features are derived from the region according to DL modelling, which is used to recognize the pathological details with minimum complexity. The efficiency of the deep model has been tested with MRI brain scan images of tumor patients. The evaluation results show that BK-BGAN's accuracy rate is increased upto 0.24%, and the prediction loss rate is optimized upto 0.28% than other deep models involved in the comparison.

Keywords Digital Pathology · Deep learning (DL) · Metastases · Prewitt Kernel · Medical images · Pathology features

S. K. Abd (✉)
Computer Techniques Engineering, Dijlah University College, Baghdad, Iraq
e-mail: sura.khalil@duc.edu.iq

M. M. Jaber
Information Technology Department, Dijlah University College, Baghdad, Iraq
e-mail: mustafa.musa@duc.edu.iq

S. M. Ali
Computer Science Department, Dijlah University College, Baghdad, Iraq
e-mail: saif.alameri@duc.edu.iq

S. K. Abd · M. M. Jaber
Al-turath University College, Baghdad, Iraq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_55

1 Introduction

Medical imaging [1, 2] is characterized as specialized equipment and methods utilized to develop the pathological images of various parts of a patient body for various clinical reasons and research. Lots of imaging modalities are developed in past two decades, therefore, few of the imaging techniques widely available today are X-ray, CT, PET [3], and magnetic resonance imaging (MRI) [4]. Using these imaging approaches and special diagnosing equipment's, information of particular tissues or organs are retrieved, which helps in various clinical purposes such as treating diseases, support medical surgery. During pathological image analysis process, anyone of the modality based scanned images are used to identify pathological changes, physical status, and so on. Moreover, the selection of appropriate modality based images is also a significant task for obtaining required information for successful pathological analysis. Positron Emotion Tomography (PET), ultrasound, and many derivative protocols of MRI [5] (diffusion-tensor MRI and fMRI) are used to retrieve information of physical structures with respect to metabolic activities. In the field of pathological image analysis, image accusation, pathological region identification (segmentation) [6], pathological feature extraction [7], significant feature selection [8], and model training (classification) are the main research areas and each contains sub-parts as well. From the past few years, researchers are working on lots of pathological imaging applications areas such as various tumors [9], neurological disorders [10], and various chronicle diseases. Out of many diseases in pathological science, brain tumor is one of the common problems the world facing today. Glioma [11] is one of the growing research areas in the field of tumor identification [12], it commonly occurs in various region of brain and spinal cord. The gliomas is a category of brain tumors occur in glial supportive cells and surround nerve cells, which is helped them to perform actions. The tumor is categorized by genetic features and the kind of glial cell presents in tumor region. Therefore, early detection [13] and taking appropriate treatment can reduces the mortality rate significantly. Hence, in this research, the pathological region identification and disease classification approaches are combined to improve efficiency of glioma brain tumor early detection with the help of MRI imaging. The traditional DL models are facing difficulties to quantifying the mitoses and inflammation specific features, it also suppress the classifier prediction performance. To resolve this issue, edge feature information based segmentation approach is incorporated for accurate identification of pathological region's (lesion, mitoses, and inflammation) pixel information; it helps to increase the accuracy of classifier's prediction performance. It has been achieved by filtering edges using Prewitt kernel approach and identifying pathological regions pixels using adaptive k-means based segmentation. In this research not only contribute to increase pathological region identification, it also contribute to improve the classifier efficiency by optimizing the network functionalities. Therefore, the buffer-fly optimizer algorithm (BOA) has been utilized to optimize the functionalities of GAN classifier.

Initially, the Prewitt kernel and adaptive k-means based automatic pathological region segmentation is performed. Next Pathological feature information is extracted using statistical approaches. The irrelevant features are extracted with f-static approach before training the model. Finally, generative adversarial network has been optimized to improve the tumor detection performance.

The research has been organized in the subsequent order, Sect. 1 discusses introduction, Sect. 2 discusses the literature review, Sect. 3 discusses the methodologies, and Sect. 4 describes the efficiency of the PK-BGAN model. Section 5 discusses the conclusion of the results finding, and finally with the reference details at the end.

2 Literature Review

This section discusses the related study on various pathological image analysis technique and related methodologies. A study by Zheng et al. [14], utilized an adaptive k-means approach for accurate segmentation of input image. Initially, it transforms the color space of image into lab color space and luminance component value is set to particular value to reduce the effect of light on image. Finally, equivalent relationship between k values and number of connection domains are used to segment the image after setting threshold. Jia and Chen [15], presents a fully automatic heterogeneous segmentation has been utilized in support vector machine for brain tumor identification. Moreover, image features are extracted using statistical feature extraction approaches to improve the classification accuracy of deep learning model. In [16] presents an optimization scheme to detect edges in an image. In this research, Prewitt kernel approach and de-noising of BM3D are concatenated to solve edge misjudgment issue in edge detection algorithm, it. This research [17], frequency domain and entropy based features are fused to create feature vector. Eight different features are extracted for each EEG signals and F-statistic values are used to rank the features across all the channels. The selected features are taken as input to SVM classifier for epileptic seizure detection. A research [18], introduced a novel deep Generative Adversarial Network (GAN) model for spatio-temporal ST precise prediction by implicitly learning the features. This model incorporates auto-encoder and decoder in Deep GAN structure. It contains two major moles, one is a deep ST feature learning model to handle ST correlation and stochastic elements and the fusion model is incorporated to external factors. The results shows that the DGAN approach is achieved higher accuracy in handling the ST data than 14 mainstream baseline methods. This research [19], introduces an optimization approach that mimics food searching behavior of butterflies, which is find out the location of nectar or mating partner with the help of sense of smell. Moreover, the butterfly optimizer algorithm (BOA) has been utilized for feature selection [20] to improve classification accuracy.

In [21], introduces a fuzzy c-means (FCM) clustering algorithm based brain tumor segmentation and extraction of pathological regions from MRI images to train the convolution neural network CNN based classification model training. The experimental result shows that the FCM based convolution approach has been obtained 97.87% accuracy rate. This study [22], developed a classification framework for brain tumor detection by combing statistical feature extraction with back-propagation neural network (BNN) classifier. Moreover, the feature set is created by combing 2D Gabor filter, 2D discrete wavelet transform (DWT), and statistical texture feature extraction methods. The classification model obtained maximum accuracy rate for various brain tumor datasets such as meningioma (91.9%), glioma (96.29%), and Pituitary tumor (95.66%). This study [23], the tumor classification has been done using super-pixel identification. It uses super-pixel identification method in fuzzy c-means to segmentation tumor region. Moreover, the image preprocessing has been done by anisotropic diffusion and dynamic stochastic resonance based enhancement techniques. This approach obtained maximum accuracy rate (97.94% and 98.67%) for glioma detection for ACRIN and BRATS dataset. In [24], a deep CNN model is utilized to identify various brain tumors and appropriate classes. This network architecture is configured with three convolution, three Relu, one normalization, and three max-pooling layers. It achieves maximum of 98.7% accuracy rate in tumor detection. This study [25], an automatic tumor detection framework has been introduced. In this, ResNet-101 models takes MRI image slices as input to squeeze and excitation. It uses intensity smoothing and zero centering approaches for preprocessing. The method obtained maximum of 93.83% average accuracy rate.

This study [26], presented an extreme learning model for tumor detection. It utilizes triangular fuzzy median filter to enhancement and Gabor filter is utilized for texture feature information extraction. The extreme model designed using single feed forward network. The method has been achieved 99% accuracy. In [27], four separately trained SegNet models are integrated by post-processing to attain maximum feature maps. The feature values are taken to train the decision tree (DT) for classification and this approach achieves 85% f-score for whole tumor detection. This study [28], presents an automatic deep multi scale 3 dimensional (3D) CNN model for glioma classification from high grade glioma (HGG) T1 Gado MRI sequence and low grade glioma (LGG). Moreover, the method also adapts the intensity smoothing and adaptive contrast enhancement techniques. The deep model achieved 96.49% overall accuracy. The study [29], demonstrated the feasibility of employing deep learning methods to assist brain surgery. The deep automatic segmentation approach contains two connected parts such as encoder and decoder relationships. In encoding, the deep CNN extracts spatial feature information and semantic map is inserted into the decoder to get full resolution probability. The encoding has been performed using Residual network ResNet, Dense CNN, and NASNet. The maximum dice and hausdroff distance score are obtained by the deep models are 0.81–0.84 and 9.8–19.7 respectively. This study [30], comparative analysis to investigated the performance of nine deep CNN

approaches for tumor detection. The evaluation results shows that the VGG-16 model is achieved maximum of 98.71% accuracy.

According to the previous study, most of the researches are utilized various deep CNN models for cancer identifications. From the various researcher opinions, guidelines, in this work pixels using adaptive k-means based segmentation and deep learning model is utilized to analyzing the pathological information. The detailed working process of introduced approach is explained in the following section.

3 Methodologies

This section discusses the various methods and its functionalities, which are utilized in this research. The traditional Deep models difficulties in quantifying the mitoses and inflammation specific features. Moreover, the irrelevant features and segmented regions reduce the performance of digital pathology detection. Therefore, an optimized deep model is introduced in this research, which contains various recent methodologies to investigate and solve each problem exclusively. The combined Edge feature detection and Segmentation technique are utilized to quantifying the mitoses and inflammation specific features. Moreover, feature extraction and selection techniques are utilized to obtain only essential feature information. This research not only contributes to fulfill the research gaps in previous studies, it also focused on to improve the classification performance of deep GAN classifier by incorporating buffer-fly optimization algorithm and Prewitt kernel.

Figure 1 illustrates the brain tumor detection process of PK-BGAN based classification model. Initially, the MRI brain image had been collected from various database sources. The classification system contains two distinguish models to train the network such as generator and discriminator. Discriminator directly takes real images as input to train the model but the generator randomly selects image samples and applies noise on it (to generate fake image); before providing input to discriminator. The preprocessing phase contains four stages image resizing, combined Prewitt kernel edge filtering and adaptive k-means based pathological regions segmentation, statistical measure based texture feature extraction, and f-static based feature selection. After preprocessing, the consolidated features are feed into deep fully connected layers of deep generative adversarial network to train the model. The discriminator calculates probability to know the prediction loss to decide whether the given MRI image contains true AD stages or generator's fake image. Whenever the discriminator encounters fake image of generator (based on loss), the GAN nodes are back-propagated and optimize networks weight with the help of butterfly optimizer to retrain the discriminator model and as well as generator until the model state activation condition satisfy to train real tumor image. The methodologies of PK-BGAN and their functionalities are discussed in this section.

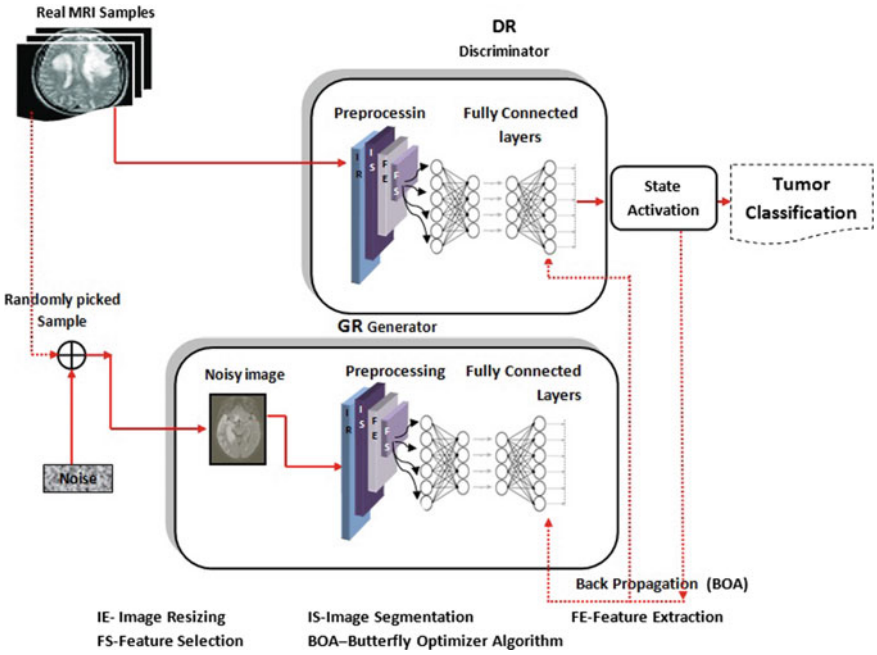


Fig. 1 Workflow of the PK-BGAN based pathological image analysis

3.1 Image Acquisition

Generally, the PK-BGAN model is designed to analysis and classifies the pathological features of medical MRI images. Generally, the 3T structural T1 weighted MRI scans images are obtained using 1.5T or 3T scanners. However in this research, the MRI baseline images are utilized to train the classifier, which are taken from various publicly available datasets such as, BRATS [31], CE-MRI [32], ACRIN 6684 [33]. The BraTs dataset contains MRI brain images of 336 heterogeneous glioma patients, including 256 HGG, and 76 LGG. CE-MRI dataset contains three types of tumors such as glioma, meningioma, and pituitary. The datasets contains 3064 MRI brain images of 233 patients. ACRIN-FMISO-brain (ACRIN 6684) dataset images are scanned from 30 patients in different angels. From the overall MRI image collections, 80% images are taken to train the PK-BGAN model and 20% images are utilized to test the efficiency of the model.

Preprocessing of the images the pathological grayscale images are resized into 300 X 300 matrixes, which helped to reduce the memory utilization and increase the performance of the classifier as well. The resized MRI images are utilized as input for the subsequent analysis process.

Figure 2 contains the sample resized MRI brain images of glioma tumor. The images contain lesion regions, mitoses regions, and inflammation regions.

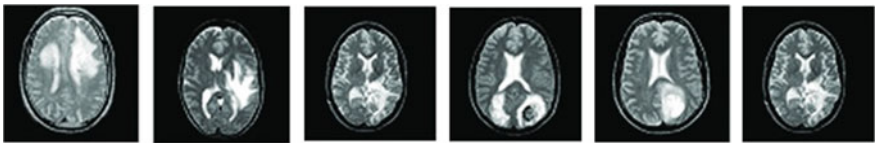


Fig. 2 Sample resized MRI brain tumor images

3.2 Prewitt Kernel

In this research, the Prewitt kernel is incorporated with segmentation approach to filter and detect the edges features of pathological regions for a given input medical image. Prewitt is a filtering operator, commonly used for edge feature detection. It works similar to Sobel operator, in that, it uses two 3×3 kernels. It is also known as first order differential operator for edge detection. The edge detection is performed using the gray variations among upper and lower pixel points, left and right adjustment points. The edge is detected by extreme value at edge points, and it also removes few pseudo-edges to smooth noise. In this, horizontal and vertical edges are detected by two directional templates. One calculates the variations in horizontal direction, and another one calculates the variations in the vertical direction. The two kernels are convolved with the original image to calculate the approximation of derivatives. Each point contains the horizontal D_h and vertical derivative D_v approximations with two images, which are represented in the following manner,

$$D_h = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} * Img* \tag{1}$$

$$D_v = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} * Img \tag{2}$$

In Eqs. (1) and (2) *indicate the 2 dimensional convolution operation. The source image is denoted as *Img*. Since the Prewitt kernel is divided as the product of an averaging and differential kernel, it performs gradient with smoothing.

$$\begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} [\,+1 \quad 0 \quad -1\,] \tag{3}$$

Therefore, it is a separable filter. The D_h is rewritten as in Eq. (3). The x and y-coordinates are defined as increasing in the left direction and increasing up direction respectively.

$$D = \sqrt{D_h^2 + D_v^2} \tag{4}$$

The resulting gradient approximations is concatenated to produce the gradient magnitude, at each point of image using Eq. (4).

$$\Theta = a\tan 2(D_h, D_v) \tag{5}$$

The derivative to calculate gradient direction is denoted in Eq. (5). For example Θ is 0 for a vertical edge which is darker on the right side.

$$D_h = \frac{\partial f(h,v)}{\partial_h} = f(h+1,v) - f(h,v) \tag{6}$$

$$D_v = \frac{\partial f(h,v)}{\partial_v} = f(h+1,v) - f(h,v) \tag{7}$$

The derivatives of gradient's definition is given in Eqs. (6) and (7).

$$grad(f(h,v)) = [D_h, D_v]^T = \left[\frac{\delta f}{\delta h}, \frac{\delta f}{\delta v}\right] \tag{8}$$

The gradient is a vector, the Eq. (8) indicates the gradient at (h, v) coordinate points.

$$|grad(f(h,v))| = \sqrt{D_h^2 + D_v^2} \tag{9}$$

The Eq. (9) indicates the size of the gradient vector.

$$\delta(h,v) = \arctan\left[\frac{D_h}{D_v}\right] \tag{10}$$

The Eq. (10) indicates the direction of the gradient. The Prewitt kernel is used to filter the edge feature from the MRI brain image. The edge filter is detects edge features of MRI brain images and smoothen the pathological image noise as well. The edge feature is combined with segmentation approach for accurate pathological region identification.

3.3 K Means Based Image Segmentation

The pathological region segmentation is a one of the essential step in medical image analysis. It is not only used for extracting pixel information from tissue images, the visualization of the segmented pathological region helps during medical treatment

and brain surgical process as well. Generally, a convolution feature of any Convolution neural network CNN model is failed to identify pathological region before train the model. It may lead misclassification as a result. Therefore, in this research, the adaptive k-means approach is attached with Prewitt kernel to improve the accuracy of feature identification. Moreover, pathological regions of medical images are segmented before model training. The steps for automatic pathological segmentations are derived in this section.

Initialize the k value to initialize seeds from the image x,

$$d\left(x_a, p_i^t\right) = \sqrt{\sum_{i=1}^{np} \left(\mathrm{p_i} - x_a\right)^2} \tag{11}$$

$$Seg_i^t = \left\{ x_a : \left\| x_a - p_i^t \right\|^2 \leq \left\| x_a - p_{np}^t \right\|^2 \forall np, \quad 1 \leq np \leq k \right\}, \tag{12}$$

$$p_i^{t+1} = \frac{1}{|\, Seg_k^t \,|} \sum_{x_{np \in Seg_i^t}} x_{np} \tag{13}$$

where $k = p_1, p_2, \ldots, p_k$ where p denotes means. Assign each pixel x_{ij} values into at least one of the nearest seed group based on the smaller distance value. The Euclidean distanced calculated by the Eq. (11). In Eq. (12) each pixel should assigned in at least one segment. Let 'a' is selected pixel point to calculate distance, d denotes the distance vector and the 'np' represents number of pixels. In update step, update centroid (p_i^{t+1}) value is calculated using Eq. (13) in each iteration. In stopping criteria repeat the Eqs. (12) and (13) till two iterations get the same segment results then stop the process. Finally the segmented regions of the images are taken as input for further processing steps such as edge detection, feature extraction, feature selection, and classification.

Figure 3 shows the sample segmented lesion, mitoses, and inflammation regions using adaptive k-means approach. In this figure, the tumor regions are segmented and applied mask in red color. The segmented pixels regions are taken for pathological texture features information extraction.

3.4 Pathological Statistical Feature Extraction

To improve the prediction performance of the classification model, few statistical features extraction techniques are also included. The mathematical derivations of the statistical features are,

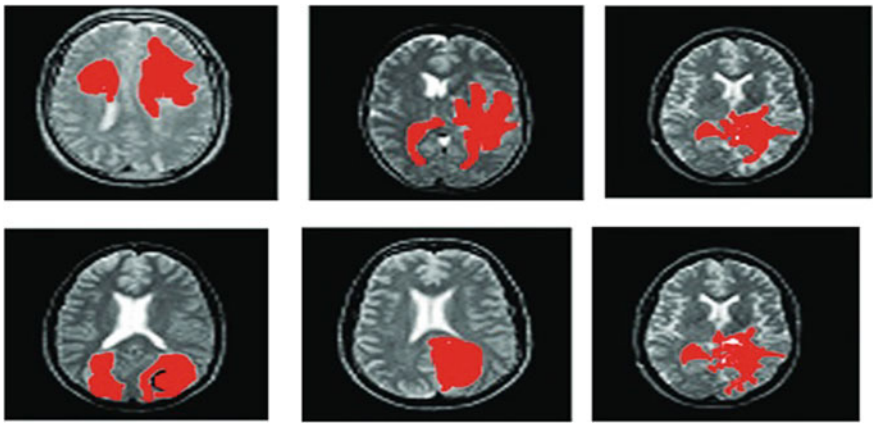


Fig. 3 Segmented pathological regions

$$\text{Mean of signal, } \mu = \frac{\sum VS}{ns} \tag{14}$$

where in Eq. (14) the μ is mean of the segmented pathological region, *vs* denotes the pixel value of at particular instant and *ns* indicates the number of pixels in a segmented image region. The Eq. (4) is used extract the mean value of segmented region.

$$\text{Coefficient of variance, } cv = \frac{\sqrt{\frac{\sum (vs-\mu)^2}{ns-1}}}{\mu} \tag{15}$$

The *cv* in Eq. (15) is used to calculate coefficient of variance values for pathological region.

$$\text{Standard deviation } \sigma = \sqrt{\frac{\sum (vs-\mu)^2}{ns-1}} \tag{16}$$

Correlation

$$Cr = \frac{\sum_{r=1}^{ndg} \sum_{c=1}^{ndg} f(r,c)(r,c) - \mu_r \mu_c}{\sigma_r \sigma_c} \tag{17}$$

where in Eq. (17) *ndg* denotes the number of discrete gray value, the *Cr* is used to calculate the correlation feature of an image. *r* and *c* indicates the horizontal and vertical special domain value respectively. The symbol σ denotes the standard

deviation of a segmented region in Eq. (16). After the statistical feature extraction feature vectors are construed to train the deep model.

3.5 Feature Selection

It is essential to identify and discard irrelevant feature and discarding irrelevant feature from the feature vector before training deep model to reduce wrong prediction. In this research, F-static value is used for feature ranking and selection process of extracted pathological features. It measures the ratio of between class variances and the mathematical expression is defined as,

$$Fs = \frac{\sum_{v=1}^{V} (\overline{\mu}_v - \overline{\mu})^2 / (V-1)}{\sum_{v=1}^{V} \sum_{j=1}^{n_v} (\overline{\mu}_v - \overline{\mu})^2 / (N-v)} \tag{18}$$

where in Eq. (18) n_v denotes the number of samples in v-th variance, $\overline{\mu}$ denotes the mean of all the observations and $\overline{\mu}_v$ indicates the mean of v-th variance. In this, large F-static value indicates the significant feature. The threshold φ_k is determined by averaging F-static values among the remaining channels $\complement_i$ and the Eq. (18) is used for feature ranking.

$$\varphi_k = 1/\complement_k \sum_{c=1}^{\complement_i} Fs_k^C \tag{19}$$

where in Eq. (19), Fs_k^C represents the F-static values of k-th feature from channel c. It ranks k-th feature across all the channels and select subsets of k-th features with high Fs_k^C. The features have low Fs_k^C value are discarded. The Eq. (19) is used to estimate the threshold value to select best features. The selected significant tumor's pathological features are taken for classification to improve the prediction performance.

3.6 Deep Generative Adversarial Network

In this research, Deep Generative Adversarial network has been used to detect brain tumor by using pathological features. GAN is an architecture in which two opposite networks compete with each other to generate desired output image. These two networks are named as generative network and discriminative network. The main goal of the optimized Generator GR is to generate images similar to samples which are generated from true image data distribution, while Discriminator DR's purpose

is to differentiate between image sample generated from GR (fake) and sample generated from the true data distribution (real).

DR assign higher probability to real MRI images and lower probability to the MRI image samples generated by generator GR, where GAN training simultaneously keep trying to remove the generated MRI samples towards the real data manifolds using the gradient information given by the DR.

3.7 Binary Cross Entropy Loss

The MRI image s generated from the true image distribution, q_{data}, noise vector z comes from a prior distribution q_z and q_g is the GR's distribution over image s. The GR takes hidden vector z as input and outputs a sample GR(z) with the purpose of bringing GR(z) as close as possible to GR(s). DR is simply a classifier in which $DR(s) = 1 \quad if\ s \sim q_{data}$ and $DR(s) = 0 \quad if\ s \sim q_g$. A min max objective is used to train both GR and DR models jointly solving as follows,

$$\min_{\theta_{GR}} \max_{\theta_{DR}} V(GR, DR) = \min_{GR} \max_{DR} E_{s \sim q_{data}}[\log DR(s)] + E_{z \sim q_z}[\log(1 - DR(GR(z)))] \tag{20}$$

where the binary cross entropy function $V(GR, DR)$ is utilized in binary classification problems. Both GR and DR are trained by back-propagating the loss in Eq. (20) through their respective models update parameters. Due to the two different objectives in Eq. (20), the update rule is defined as follows:

$$\left\{\theta_{DR}^{t+1}, \theta_{GR}^{t+1}\right\} \leftarrow \begin{cases} update\ if\ DR(x) predicts\ wrong \\ update\ if\ DR(GR(z)) predicts\ wrong \\ update\ if\ DR(GR(z)) predicts\ correct \end{cases} \tag{21}$$

where in Eq. (21) θ_{DR}^{t+1} and θ_{GR}^{t+1} are the parameters of DR and GR respectively, and t is the iteration number. In recent times, DGAN is developed as an innovative technique like adversarial loss for learning generative models of complex problems. In this research, the fitness/objective function of global search space in butterfly optimization algorithm has been utilized to calculate the loss value for both generator and discriminator network.

3.8 Adversarial Loss

The adversarial loss of $DR(L_{GAN}^{DR})$ is derived as follows,

$$L_{GAN}^{DR} = \|DR(y_{real}) - 1\|_2^2 + \left\|DR\left(y_{fake}\right) - 0\right\|_2^2, \tag{22}$$

In Eq. (22), the adversarial losses are used to find the stability between GR and DR during the adversarial training. GR tries to minimize the $DR\left(y_{fake}\right)$ with true labels to minimize the GR loss L_{GAN}^{GR}. GR's objective is to create real lookalike sample images to DR, therefore, to minimize the GR loss L_{GAN}^{GR}, GR tries to reduce the difference between $DR\left(y_{fake}\right)$ and true label.

$$L_{GAN}^{GR} = \left\|DR\left(y_{fake}\right) - 1\right\|_2^2 \tag{23}$$

The overall objective of the DGAN is derived as follows,

$$L_{DGAN} = L_{GAN}^{GR} + L_{GAN}^{DR} \tag{24}$$

where the Eq. (23) is used to reduce the loss difference between true label and DR's fake labels, and the Eq. (24) is used to calculate the overall loss for DGAN, which is used replaces the traditional binary classification loss values of GR and DR in Eq. (20). Moreover, both GR model and DR model loss values in Eq. (22) has been calculated with the help of (butterfly global search space solution based fitness/objective function in Eq. (27)) BOA through their respective models updates parameters.

3.9 BOA Optimizer

Instead of its remarkable success in generating realistic images, training a GAN is really a challenging task due to various reasons, such as difficulty in convergence, model collapse and vanishing gradients. In this research, the fitness/objective function of global search space in BOA has been Instead of its remarkable success in generating realistic images.

The characteristics of butterflies are defined by two main issues: the variance of stimulus intensity SI and formulation of recognized magnitude of the fragrance/smell *fr* (how stronger the fragrance is recognized by other butterflies). For simplicity, *SI* of a butterfly is connected with the encoded objective function. The smell/fragrance is formulated as a function of physical intensity of stimulus in BOA using this concept, which is derived in following derivation,

$$fr = smSI^e \tag{25}$$

In Eq. (25), *sm* is sensory modality, and *e* is the power of exponent dependent on modality, which reports the differing degree of absorption. The range of *e* and *sm* is taken from $[0, 1]$. The BOA includes three stages such as, Initialization, Iteration,

and Finale phase. The algorithm first estimates the fitness value for all the butterflies on various position in a solution space. In this algorithm have two important stages such as, global and local search.

$$b_u^{it+1} = b_u^{it} + \left(rnd^2 * h^* - b_u^{it}\right) * fr_u \tag{26}$$

The Eq. (26) denotes the derivative of global search stage and the butterfly moves towards the fittest butterfly or solution h^*. b_u^{it} is the solution vector b_u for u-th butterfly in iteration number it. The rnd is a random number in [0, 1] and the smell of u-th butterfly is derived by fr_u. The h^* denotes the current best solution among all the solution in current iteration.

$$b_u^{it+1} = b_u^{it} + \left(rnd^2 * b_v^{it} - b_w^{it}\right) * fr_u \tag{27}$$

The Eq. (27) is denoted as local search phase of the BOA. Where b_v^{it} and b_w^{it} are v-th and w-th butterflies from solution space. If b_v^{it} and b_w^{it} butterflies belongs to same swarm and rnd is a random number between the range [0, 1] then the Eq. (27) becomes a local random walk. A switch probability sp in BOA helps to switch between common global to local searches. This iteration has been continued until the stopping criteria are not match. In this research, both GR model and DR model loss values in Eq. (22) has been calculated with the help of (butterfly global search space solution based fitness/objective function in Eq. (27)) BOA through their respective models updates parameters.

3.9.1 ReLU-Activation Function

ReLU is less computational complex than sigmoid and tangent activation function and helps to avoid vanishing gradient problem as well. Therefore, ReLU is utilized for state activation function in the PK-BGAN model. The Rectified Linear Unit ReLU function $max(0, z)$ and derivative $ReLU'$ is denoted as follows,

$$ReLU(z) = \begin{cases} z & z > 0 \\ 0 & z \leq 0 \end{cases} \quad \forall_z = -1, \ldots 1 \tag{28}$$

$$ReLU'(z) = \begin{cases} 1 & z > 0 \\ 0 & z < 0 \end{cases} \quad \forall_z = -1, \ldots 1 \tag{29}$$

The state activation conditions of ReLU has been given in Eqs. (28) and (29). During the model training, the prediction loss of each nodes and hyper parameters values has been updated as stage by stage. The state activation functionalities of the framework are performed by ReLU. The ReLU function activates current state output to train the model when the parameters value reaches between −1 and 1.

The objective function is one of the most fundamental components of any deep learning model; it provides the basic formal specification of a problem. In this research, the global search phase objective function of buffer-fly optimizer algorithm is incorporated generative adversarial network model to improve the prediction performance of GAN.

3.9.2 BGAN Algorithm

Input : Feature vector of MRI brain images

α-learming rate, m-batch size , itr– number of iterations

Initialize: $w, \alpha = 0.0002, m = 30, itr = 10$

$(y_{real}) = \{xx^{(1)}, xx^{(2)}, \ldots, xx^{m}\}$ // from prior xx noisy image (feature vector)

$(y_{fake}) = \{x^{(1)}, x^{(2)}, \ldots, x^{m}\}$ // from prior x discriminator actual image (feature vector)

While$max\ (0, y)$**do**

For t= 1 to itr

$L_{GAN}^{DR} = \|DR(y_{real}) - 1\|_2^2 + \|GR(y_{fake}) - 0\|_2^2$, // calculate discriminator loss

$w = w + \alpha * BOA(\ L_{GAN}^{DR})$ // update discriminator loss using butterfly optimizer

Endfor

$L_{GAN}^{GR} = \|DR(y_{fake}) - 1\|_2^2$ // calculate generator loss

$$L_{DGAN} = L_{GAN}^{GR} + L_{GAN}^{DR}$$

$O_t = L_{DGAN} - \alpha *$ BOA (L_{GAN}^{GR})// update generator loss using butterfly optimizer

Endwhile

O_t // train the predicted value

Output: predicted disease class

The Feature vector of MRI brain images has been taken as input to train the model to classify diseases. Initially, it initializes parameters values and compute ReLU activation. Calculate loss for discriminator and generator model and update the weight by using BOA. Continue this process until it reached global convergence value. Whenever the convergence values reaches to ReLU activation threshold train the model to classify disease class. Continue this process for each image in a batch. Finally, the model classifies tumor class for each batch images. The subsequent section discusses the evolution results and efficiency of the PK-BGAN model.

4 Result and Discussions

This section describes the performance analysis of PK-BGAN classifier in tumor prediction, the excellence of the tumor identification is determined with various evaluation metrics [34, 35] such as accuracy, sensitivity, and f-score. The PK-BGAN

model has been implemented in python's keras API. The pathological image analysis system performance has been compared with recent accuracy wise best classifiers such as GAN [18], CNN [24], and GG-16 [30].

$$\text{Accuracy rate, } Acc = \frac{TrPo + TrNe}{TrPo + TrNe + FaPo + FaNe} \tag{30}$$

$$\text{Sensitivity (Recall) rate, } Sen/Rec = \frac{TrPo}{TrPo + FaPo} \tag{31}$$

$$\text{Precision rate, } Pre = \frac{TrPo}{TrPo + FaPo} \tag{32}$$

$$\text{F-score } f_1 = \frac{2 * Pre * Rec}{Pre + Rec} \tag{33}$$

where In Eq. (30) *TrNe* represents true negative rate, *TrPo* represents the true positive rate, *FaPo* indicates false positive rate, and the *FaNe* indicates false-negative rate. The Eqs. (30–33) are used to estimate the efficiency of tumor identification using MRI brain imaging.

The Table 1 contains the average loss value of CNN, VGG-16, GAN, BK-BGAN models for tumor detection. Among these four deep learning models, a BK-BGAN model has been obtained less loss rate (0.04%). It clearly proves that the BK-BGAN loss rate is reduced minimum of 0.28% than other deep models.

Figure 4, demonstrates the comparison loss value of predicted BK-BGAN models and actual values for tumor detection. It provides evidence that the BK-BGAN model is attained less loss rate (0.04%) in tumor classification with the help of buffer-fly optimizer's global search space fitness function $x_i^{t+1} = x_i^t + \left(r^2 \times g^* - x_i^t\right) \times f_i$.

The Fig. 5 demonstrates the comparison prediction loss value of BK-BGAN, GAN, VGG-16, and CNN models for tumor detection. It provides evidence that the adaptive k means based pathological region segmentation $Seg_i^t = \left\{ x_a : \left\| x_a - p_i^t \right\|^2 \leq \left\| x_a - p_{np}^t \right\|^2 \forall np, 1 \leq np \leq k \right\}$ approach has been identified and grouped exactly correct pixels.

Table 1 Average loss rate

Classifier	Loss (%)
CNN	2.46
VGG-16	2.41
GAN	0.32
BK-BGAN	0.04

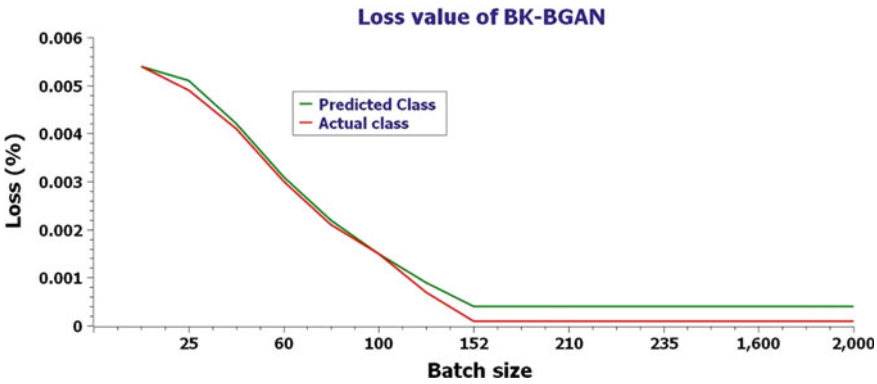


Fig. 4 Loss of BK-BGAN model for CE-MRI dataset

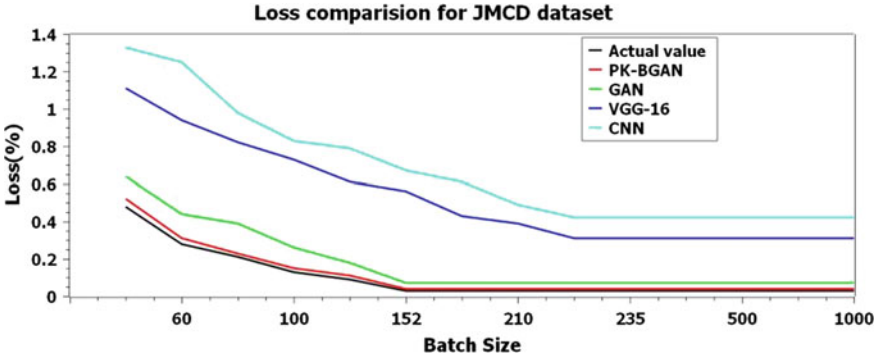


Fig. 5 Loss comparison of BK-BGAN, GAN, VGG-16, CNN model for ACRIN dataset

The Fig. 6 demonstrates the comparison prediction loss value of BK-BGAN, GAN, VGG-16, and CNN models for tumor detection. It provides evidence that the derivative of prewitt kernel's gradient direction (horizontal and vertical)$\Theta = a\tan 2\left(G_y, G_x\right)$ is filtered the exact pathological edges. It helps to reduce the loss of BK-BGAN classifier than comparison approaches.

The Table 2 contains the average accuracy, f1-score, and sensitivity value of CNN, VGG-16, GAN, and BK-BGAN models for tumor detection. Among these four deep learning models, a BK-BGAN model has been obtained maximum accuracy rate (0.24%), sensitivity rate (0.13%), f1-score value (0.10%) than comparison methods.

The Table 3 contains the average accuracy value of CNN, VGG-16, GAN, BK-BGAN models for tumor detection. Among these four deep learning models, a BK-BGAN model has been obtained maximum accuracy rate (99.98%). It clearly proves that the BK-BGAN accuracy rate is increased (0.24%) than comparison deep models.

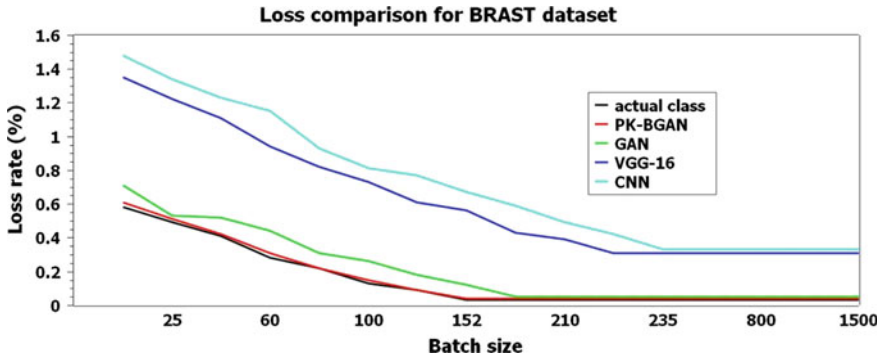


Fig. 6 Loss comparison of BK-BGAN, GAN, VGG-16, CNN model for BRAST dataset

The Fig. 7 demonstrates the comparison prediction accuracy of BK-BGAN, GAN, VGG-16, and CNN models for pathological region based tumor detection. It provides evidence that the F-static approach estimates exact thresholds ($\varphi_k = 1/\complement_k \sum_{c=1}^{\complement_i} Fs_k^C$) to select significant pathological features to train the model. The feature helps to increase the accuracy of BK-BGAN classifier than comparison approaches.

The Fig. 8 shows the comparison f-score value for BK-BGAN, GAN, VGG-16, and CNN models in pathological region based tumor detection. It provides evidence that the combined feature of the twin network's (discriminator and generator) simultaneous image analysis and global search space fitness function of butterfly optimizer ($w = w + \alpha * BOA\left(L_{GAN}^{DR}\right)$ and $O_t = L_{DGAN} - \alpha *$ BOA (L_{GAN}^{GR})) can helps BK-BGAN model to achieve maximum accuracy rate than other deep CNN based tumor classification.

The Fig. 9 shows the comparison f-score value for BK-BGAN, GAN, VGG-16, and CNN models in pathological region based tumor detection. It provides evidence that the adaptive k means based pathological region segmentation $Seg_i^t = \left\{ x_a : \left\| x_a - p_i^t \right\|^2 \leq \left\| x_a - p_{np}^t \right\|^2 \forall np, 1 \leq np \leq k \right\}$ approach has been identified and grouped exactly correct lesion pixels. It helps BK-BGAN model to achieve maximum sensitivity rate than other deep CNN based tumor classification. The overall result proves the efficiency of the specific problem oriented functionalities of the PK-BGAN framework. First, the exact lesion and inflammation region's pixels identification using the prewitt kernel's based edge filtering and adaptive segmentation helps to improve accuracy. Next, the global search space fitness function of BOA in GAN facilitates to optimize the network loss rate during the tumor classifications. These functionalities in the classification framework influence to achieve better accuracy rate than comparison deep models. The overall evaluation

Table 2 Summery of average accuracy, f1-score and sensitivity value

Batch size	Accuracy (%)				F1-score (%)				Sensitivity (%)			
	BK-BGAN	GAN	VGG-16	CNN	BK-BGAN	GAN	VGG-16	CNN	BK-BGAN	GAN	VGG-16	CNN
60	**99.98**	99.31	98.71	98.7	99.58	99.26	98.67	98.45	**99.88**	99.65	98.56	**98.89**
100	99.92	99.24	98.02	98.64	**99.76**	99.54	98.56	98.54	99.65	99.34	98.65	98.77
152	99.81	99.66	**98.96**	98.56	99.33	99.45	98.44	98.35	99.34	**99.75**	98.66	98.43
210	99.84	**99.74**	98.84	98.44	99.34	**99.66**	98.32	98.57	99.43	99.34	**98.77**	98.76
235	99.77	99.65	98.77	**98.79**	99.53	99.54	98.34	**98.69**	99.55	99.32	98.23	98.67
500	99.34	99.21	98.34	98.46	99.65	99.51	**98.76**	98.65	99.67	99.45	98.66	98.55

Bold shows the highest result

Table 3 Average accuracy value

Classifier	Average accuracy (%)
CNN	98.76
VGG-16	98.96
GAN	99.74
BK-BGAN	99.98

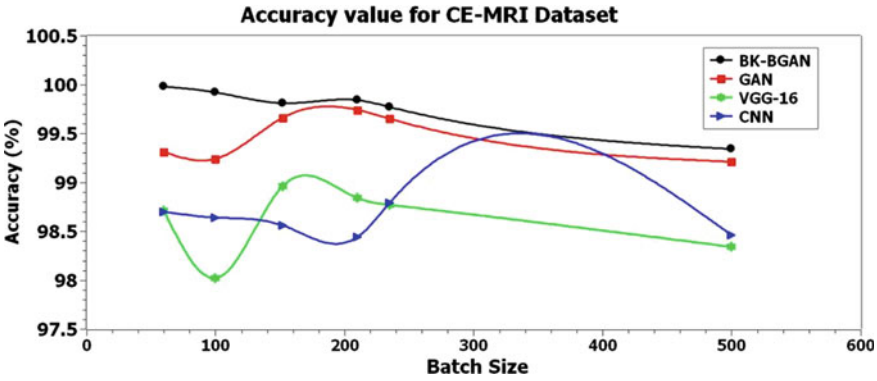


Fig. 7 Accuracy comparison of BK-BGAN, GAN, VGG-16, CNN model for CE-MRI dataset

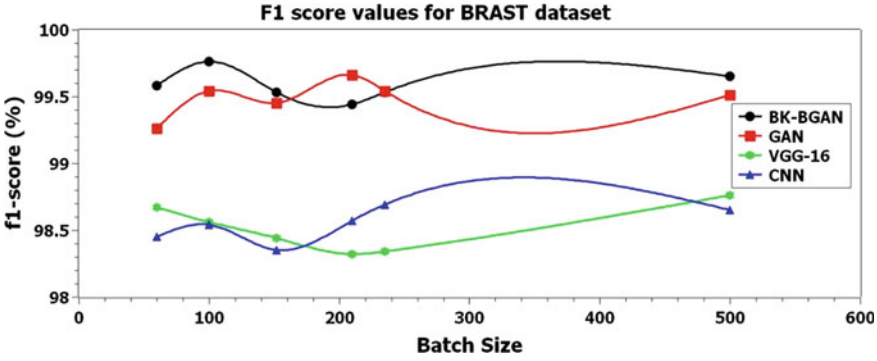


Fig. 8 F1 score value comparison of BK-BGAN, GAN, VGG-16, CNN model for BRAST dataset

results shows that BK-BGAN's accuracy rate is increased up to 0.24% and prediction loss rate is optimized upto 0.28% than other comparison deep models. The subsequent section discusses the conclusion of the research findings.

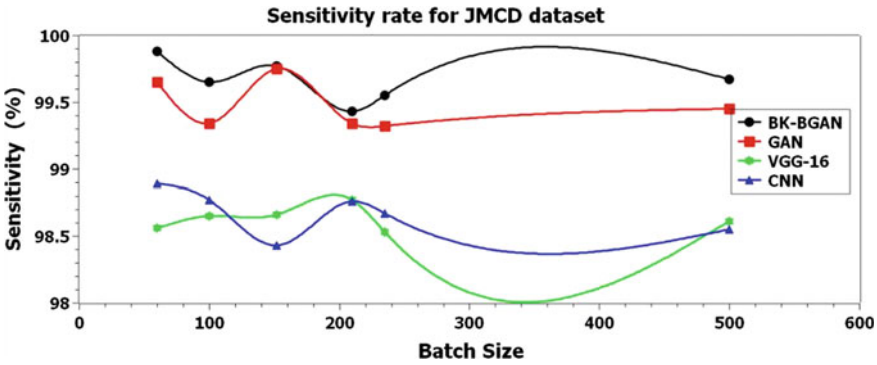


Fig. 9 Sensitivity comparison of BK-BGAN, GAN, VGG-16, CNN model for ACRIN dataset

5 Conclusion

Hence, the evaluation reports in results and discussion section show that the PK-BGAN classifier based tumor framework has been performed efficiently with all the evolution metrics than comparison deep models. It clearly shows that the new sample training feature of BOA in GAN helps to reduce the loss rate upto (0.28%) of PK-BGAN model. Moreover, the combined features of adaptive k-means and prewitt kernel based exact lesion and inflammation region's pixels information identification helps to increase the accuracy rate (0.24%). The overall evaluation results show that combined problem specific functionalities of the BK-BGAN based tumor classification model has been obtained extreme best accuracy (99.98%) and loss rate (0.04%) than comparison deep models. Therefore, the research is concluded that the BK-BGAN model is outperforms in diagnosing tumor. Furthermore, in future research the efficiency of the PK-BGAN model has been tested with other pathological MRI images of Dementia, lung cancer, and Parkinson's disease.

References

1. Yu Y, Lin H, Meng J, Wei X, Guo H, Zhao Z (2017) Deep transfer learning for modality classification of medical images. Information 8(3):91. https://doi.org/10.3390/info8030091
2. Zhang J, Xia Y, Xie Y, Fulham M, Feng DD (2018) Classification of medical images in the biomedical literature by jointly using deep and handcrafted visual features. IEEE J Biomed Health Inform 22(5):1521–1530. https://doi.org/10.1109/JBHI.2017.2775662
3. Lu D, Popuri K, Ding GW, Balachandar R, Beg MF (2018) Multimodal and multiscale deep neural networks for the early diagnosis of Alzheimer's disease using structural MR and FDG-PET images. Sci Rep 8(1). Art no 5697
4. Ortiz A, Munilla J, Martínez-Murcia FJ, Górriz JM, Ramírez J, the Alzheimer's Disease Neuroimaging Initiative (2017) Learning longitudinal MRI patterns by SICE and deep learning: Assessing the Alzheimer's disease progression. In: Proceedings of the annual conference on medical image understanding and analysis, 2017, pp 413–424. Springer

5. Kazemi Y, Houghten SK (2018) A deep learning pipeline to classify different stages of Alzheimer's disease from fMRI data. In: Proceedings of the IEEE conference on computational intelligence in bioinformatics and computational biology (CIBCB), pp 1–8
6. Rezaei M, Yang H, Meinel C (2020) Recurrent generative adversarial network for learning imbalanced medical image semantic segmentation. Multimed Tools Appl 79:15329–15348. https://doi.org/10.1007/s11042-019-7305-1
7. Yin B, Wang C, Abza F (2020) New brain tumor classification method based on an improved version of whale optimization algorithm. Biomed Sig Process Control 56:101728.https://doi.org/10.1016/j.bspc.2019.101728
8. Poernama AI, Soesanti I, Wahyunggoro O (2019) Feature extraction and feature selection methods in classification of brain MRI images: a review. In: 2019 International biomedical instrumentation and technology conference (IBITeC), Special region of Yogyakarta, Indonesia, 2019, pp 58–63.https://doi.org/10.1109/IBITeC46597.2019.9091724
9. Swapnil SA, Girish VS (2020) Image mining methodology for detection of brain tumor: a review. In: 2020 Fourth international conference on computing methodologies and communication (ICCMC), Erode, India, 2020, pp 232–237. https://doi.org/10.1109/ICCMC48092.2020.ICCMC-00044
10. Tyagi V (2019) A review on image classification techniques to classify neurological disorders of brain MRI. In: 2019 International conference on issues and challenges in intelligent computing techniques (ICICT), Ghaziabad, India, 2019, pp 1–4. https://doi.org/10.1109/ICICT46931.2019.8977658
11. Liang J, Lv X, Lu C et al (2020) Prognostic factors of patients with Gliomas—an analysis on 335 patients with Glioblastoma and other forms of Gliomas. BMC Cancer 20:35. https://doi.org/10.1186/s12885-019-6511-6
12. Wu W, Li D, Du J, Gao X, Gu W, Zhao F, … Yan H (2020) An intelligent diagnosis method of brain MRI tumor segmentation using deep convolutional neural network and SVM algorithm. Comput Math Meth Med:1–10.https://doi.org/10.1155/2020/6789306
13. Kabir MA (2020) Early stage brain tumor detection on MRI image using a hybrid technique. In: 2020 IEEE region 10 symposium (TENSYMP), Dhaka, Bangladesh, 2020, pp 1828–1831. https://doi.org/10.1109/TENSYMP50017.2020.9230635
14. Zheng X, Lei Q, Yao R et al (2018) Image segmentation based on adaptive K-means algorithm. J Image Video Proc 2018:68. https://doi.org/10.1186/s13640-018-0309-3
15. Jia Z, Chen D (2020) Brain tumor identification and classification of MRI images using deep learning techniques. IEEE Access 1–1.https://doi.org/10.1109/access.2020.3016319
16. Ye H, Shen B, Yan S (2018) Prewitt edge detection based on BM3D image denoising. In: 2018 IEEE 3rd advanced information technology, electronic and automation control conference (IAEAC). https://doi.org/10.1109/iaeac.2018.8577643
17. Peng G, Nourani M, Harvey J, Dave H (2020)Feature selection using F-statistic values for EEG signal analysis. In: 2020 42nd annual international conference of the IEEE engineering in medicine and biology society (EMBC), Montreal, QC, Canada, 2020, pp 5963–5966. https://doi.org/10.1109/EMBC44109.2020.9176434
18. Saxena D, Cao J (2019) D-GAN: deep generative adversarial nets for spatio-temporal prediction. Machine learning. arXiv:1907.08556
19. Arora S, Singh S (2019) Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput 23:715–734. https://doi.org/10.1007/s00500-018-3102-4
20. Arora S, Anand P (2018) Binary butterfly optimization approaches for feature selection. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2018.08.051
21. Hossain T, Shishir FS, Ashraf M, Al Nasim MA, Muhammad Shah F (2019) Brain tumor detection using convolutional neural network. In: 2019 1st international conference on advances in science, engineering and robotics technology (ICASERT), Dhaka, Bangladesh, 2019, pp 1–6. https://doi.org/10.1109/ICASERT.2019.8934561

22. Ismael MR, Abdel-Qader I (2018) Brain tumor classification via statistical features and back-propagation neural network. In: 2018 IEEE international conference on electro/information technology (EIT), Rochester, MI, 2018, pp 0252–0257. https://doi.org/10.1109/EIT.2018.8500308
23. Gupta N, Mishra S, Khanna P (2018) Glioma identification from brain MRI using superpixels and FCM clustering. In: 2018 Conference on information and communication technology (CICT), Jabalpur, India, 2018, pp 1–6. https://doi.org/10.1109/INFOCOMTECH.2018.8722405
24. Sultan HH, Salem NM, Al-Atabany W (2019) Multi-classification of brain tumor images using deep neural network. IEEE Access 7:69215–69225. https://doi.org/10.1109/ACCESS.2019.2919122
25. Ghosal P, Nandanwar L, Kanchan S, Bhadra A, Chakraborty J, Nandi D (2019) Brain tumor classification using ResNet-101 based squeeze and excitation deep neural network. In: 2019 Second international conference on advanced computational and communication paradigms (ICACCP), Gangtok, India, 2019, pp 1–6. https://doi.org/10.1109/ICACCP.2019.8882973
26. Sharif M, Amin J, Raza M et al (2020) Brain tumor detection based on extreme learning. Neural Comput Appl 32:15975–15987. https://doi.org/10.1007/s00521-019-04679-8
27. Alqazzaz S, Sun X, Yang X et al (2019) Automated brain tumor segmentation on multi-modal MR image using SegNet. Comp Visual Media 5:209–219. https://doi.org/10.1007/s41095-019-0139-y
28. Mzoughi H, Njeh I, Wali A et al (2020) Deep multi-scale 3D convolutional neural network (CNN) for MRI gliomas brain tumor classification. J Digit Imaging 33:903–915. https://doi.org/10.1007/s10278-020-00347-9
29. Zeineldin RA, Karar ME, Coburger J et al (2020) DeepSeg: deep neural network framework for automatic brain tumor segmentation using magnetic resonance FLAIR images. Int J CARS 15:909–920. https://doi.org/10.1007/s11548-020-02186-z
30. Chelghoum R, Ikhlef A, Hameurlaine A, Jacquir S (2020) Transfer learning using convolutional neural network architectures for brain tumor classification from MRI images. In: Maglogiannis I, Iliadis L, Pimenidis E (eds) Artificial intelligence applications and innovations, AIAI 2020. IFIP advances in information and communication technology, vol 583. Springer, Cham. https://doi.org/10.1007/978-3-030-49161-1_17
31. Brats dataset : https://www.med.upenn.edu/sbia/brats2017/data.html
32. CE-MRI dataset: https://figshare.com/articles/dataset/brain_tumor_dataset/1512427
33. ACRIN-FMISO-Brain(ACRIN-6684) dataset: https://wiki.cancerimagingarchive.net/pages/viewpage.action?pageId=33948305
34. Haq AU, Li J, Khan J, Memon MH, Nazir S, Ahmad S, Khan GA, Ali AA (2020) New intelligent approach for effective recognition of diabetes in the IoT e-healthcare environment. Preprints 2020, 2020020462. https://doi.org/10.20944/preprints202002.0462.v
35. Alarifi A, Alwadain A (2020) Killer heuristic optimized convolution neural network-based fall detection with wearable IoT sensor devices. Measurement:108258.https://doi.org/10.1016/j.measurement.2020.108258

Critical Analysis of Electronic Health Readiness Assessment Factors for Development of e-Health Framework in Iraq

Saif Mohammed Ali, M. A. Burhanuddin, Ibrahem T. Nather, Mustafa Musa Jaber, and Sura Khalil Abd

Abstract The assessment of e-Health readiness offers an accurate picture regarding the existing conditions, and a description of facilitated operational plans, as well as functional approaches to effectively adopt e-Health. To determine and assess readiness, indices need to be adjusted to particular conditions with regards to each country. To ensure that e-Health is applied successfully, pre-examination needs to be done pertaining the context of usage, so as to gather the real needs that ensure reaping all the possible advantages associated with this technology. The healthcare sector in Iraq still lags behind in terms of progress with the adoption rate of ICT systems. The country's current healthcare system is faced with various challenges that hinder the adoption of e-Health system, as well as the performance of assessment of healthcare professionals that would enable the enhancement of institutional performance. Consequently, the current study is aimed at identifying a list of success factors that could impact healthcare professionals when assessing e-Health readiness pertaining to the healthcare sector in Iraq. In line with this, eight experts, who belong to the medical and technical staffs from six different hospitals in Iraq, were interviewed. After which, a literature survey was conducted to confirm the relevancy of the collected factors so that all of these falls within the circumstance of healthcare. Four major dimensions (core readiness, environmental readiness, individual readiness and technological readiness) were identified, which

S. M. Ali (✉) · M. M. Jaber
Department of Information Technology, Dijlah University College, Baghdad, Iraq
e-mail: saif.alameri@duc.edu.iq

Al-turath University College, Baghdad, Iraq
e-mail: mustafa.musa@duc.edu.iq

M. A. Burhanuddin · I. T. Nather
Faculty of Information and Communication Technology, Universiti Teknikal Malaysia, Melaka, Malaysia
e-mail: burhanuddin@utem.edu.my

S. K. Abd
Computer Techniques Engineering, Dijlah University College, Baghdad, Iraq
e-mail: Sura.khalil@duc.edu.iq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_56

could cast an impact on the perspective of healthcare professionals while assessing e-Health readiness. Finally, based on these findings, a conceptual model was constructed.

Keywords e-Health · e-Health readiness · Critical factors · Iraqi healthcare · e-Health adoption · e-Health in Iraq

1 Introduction

With the help of information and communication technologies (ICTs), the collection, storing, exchanging and processing of electronic health systems can be carried out effectively. With the use of ICTs, considerable enhancements in healthcare services have been achieved in terms of safety, quality and cost-effectiveness [1]. It can also be defined as integrating and employing information and communications technologies simultaneously for healthcare [2]. As per Noor and Technology [3], enhancing access to healthcare services via ICTs is the need of the hour especially in developing countries. The enhancement of ICT and e-Health solutions provides different options that address the changing needs and allow networking and integration. According to the World Health Organization (2012), the prevalent use of e-Health technologies has escalated widely and swiftly, spreading not only across low-and medium-income countries but also globally.

E-readiness cannot be considered as the same as needs assessment in detecting problems, but can be defined as a measure for a society or establishment to be able to address pre-defined problems which are crucial for programs employing Information and Communications Technology. As per the majority of literary works and publications, the feedback towards the increasing technology gap among the developing and developed nations defines e-readiness [4, 5]. The study's key aim is to detect the critical factors that allow effective 'proposed preparedness' evaluation pertaining to institutions, so that e-Health can be employed effectively and implemented successfully. Hence, this study is an effort towards contributing new information to the body of knowledge in this field, particularly in Iraq.

2 Related Works

Al Hilfi et al. [6] emphasized that the current system in healthcare dissemination in Iraq is experiencing numerous problems that hinder the enhancement of the Electronic Health system, in addition to the evaluation of healthcare practitioners for the enhancement of institutional achievement. The researches stressed on the requirements for endeavors to be made for enhancement in the data quality, following its pathway transversing a single establishment right up to the Ministry of Health, which currently are unevenly distributed to meet the health needs of the

country. Additionally, [7] recognize the current challenges surrounding the Iraqi healthcare sector so as to offer better medical management services, caused by inappropriate use of technology that lacks a proper procedure for information sharing. Thus, the identification of the major barriers towards the availability to Electronic Health services in the public medical care fields of Iraq is the primary aim of this study.

The review on literature in this current study is primarily aimed at offering a setting related to the potential impediments in assessment the e-Health readiness. The empirical evidence found in earlier studies conducted in this area has been included in this literature review. The present trends in these new systems at a global and national level can be explained through the information present in the studies conducted previously. It is apparently clear from the aforementioned literature that the factors responsible for impacting the e-Health readiness assessment in hospitals in Iraq were essentially neglected. This dearth of adequate research offers the foundation for a research to evaluate the stakeholder e-Health readiness, and to assess the system requirements before introducing e-Health. Additionally, the literature review indicates that there are only a limited number of e-Health readiness assessment published in journals, involving developing countries. The journals provide a checklist to analyze and measure the ability of any country to be involved in e-readiness in order to achieve greater efficiency.

Queries were made on several electronic databases (PubMed, ProQuest, IEEE, Web of Science, Science Direct, and Google Scholar) with probe executed to ascertain publications on researches pertaining e-Health evaluation readiness. The ensuing query terminologies were utilized: “e-Health readiness”, “e-Health” AND “readiness”, “e-Health readiness assessment”, “e-Health readiness tools”, “e-Health readiness assessment frameworks”, “e-Health preparedness”, “e-Health” in addition to “preparedness”, “e-Health assessment tools”, “telemedicine readiness”, “telemedicine” as well as “readiness”, “telemedicine readiness assessment tools”, “m-Health readiness”, “m-health” alongside “readiness”, “EMR readiness” and “EHR readiness”.

The queries were inclusive of the title content, abstract, encompassing whole fields, in addition to free content. The inclusiveness measures entail the condition that the published work must be a full article, published in English language, published in a peer-reviewed journal, in addition to the paper having forwarded an e-Health readiness assessment framework or model. In the identification stage, the papers involved amounted to 2,358 papers (articles), where seven hundred and twenty-four 724 papers were replications. The procedure for choosing rejected the replicated papers from the query; which resulted in the generation of an inventory of 1,634 papers. During the screening stage, out of the 1,634 articles accumulated after removal of duplications, the remaining abstracts were reviewed to determine their suitability. Consequently, 1,418 papers were subsequently rejected as they irrelevant and not related to e-Health readiness assessment, and not published between 2012 and 2019. Then in the eligibility stage, a total of 216 full papers were examined, and 198 were rejected due to the content of the paper not describing the

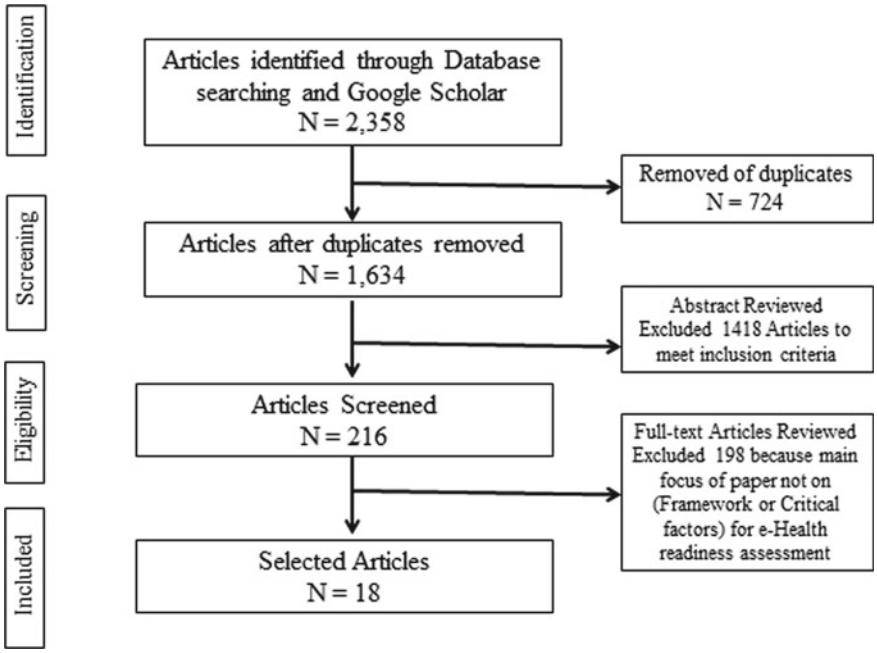


Fig. 1 PRISMA flow diagram illustrating the selection of articles procedure

critical factors or structure for evaluating e-Health readiness. Finally, during the stage for inclusion, the remaining eighteen papers constitute the foundation of the current review (see Fig. 1).

Table 1 presents a compilation of earlier studies that have been carried out under various contexts related to healthcare. Most of the factors that were recognized in the preliminary phase were supported in these studies. It is obvious from the literature that all the above factors are relevant and applicable in the context of this study.

3 Research Method

In this study, a preliminary study was performed by accounting six Iraqi hospitals to gather the key antecedents that could impact the healthcare workers to make use of e-Health technology services and implement these in their health information systems. In addition, a literature survey was conducted to collect related information pertaining to the identified factors as well as to verify if these factors can be deemed relevant with regards to other studies in the healthcare context.

To conduct the preliminary study, a qualitative research design is selected for the compilation of data within a societal milieu which is aimed at determining the

Table 1 Relevant studies pertaining e-health readiness assessment

No.	Author(s)	Individual			Environmental			Core				Technological				
		Attitudes towards	Training	Skills	Societal readiness	Government regulations	Economic	Need of change	Process of planning	Acceptability	Compatibility	Availability and affordability of hardware and software	ICT support service	Network reliability	IT resources	Privacy and security
1	Kutlu and Ozturan [8]	√														√
2	Durrani et al. [9]		√		√	√			√			√	√			
3	Rezai-Rad et al. [10]				√							√	√	√		√
4	Li et al. [11]		√		√			√				√		√	√	
5	Khoja et al. [12]		√		√	√		√	√	√		√	√			
6	Ghazisaeidi et al. [13]		√			√		√	√	√			√	√		√
7	Qureshi et al. [14]	√	√		√							√		√		
8	Buabbas [15]	√	√		√											

manner in which persons decipher as well as reason out what they have gone through [24]. A research on an empirical study on case-basis was employed [25] to perform in-depth investigation of the phenomenon within a real-life context. An interview technique was setup that employed semi-structured form, as well as eight open-ended questions that were designed to allow the respondents to provide in-depth information regarding the situation [26]. The interviewees included two technical staffs and six medical staffs who were aware of e-Health services. The target population was taken from six key hospitals in Baghdad, Iraq (Baghdad Teaching Hospital, Surgical Specialties Hospital, Private Nursing Home Hospital, Child Protection Teaching Hospital, Gastroenterology and Hepatology Centre, and Iraqi Centre for Cardiac Diseases). The aforementioned selection of hospitals was done according to their potential and ability to provide health aids and also, they comprise key hospitals employing Information Technology conveniences and equipment when offering their services. The researcher performed a thematic analysis to assess the qualitative data gathered from the interviews. As mentioned by Braun and Clarke [27], an analysis that is carried out according to themes is signified as an adaptable and uncomplicated method which can be utilized to create novel concepts through the detection, assessment, and the preparation of templates or topics, that are then organized and explained to make sense of the data.

Based on documented works in addition to the interview conducted involving the technical and medical staffs from healthcare institutions in Iraq. Among the most problematic issues impacting the evaluation of e-Health readiness in Iraqi healthcare institutions include: core readiness, environmental readiness, individual readiness and echnological readiness as represented in (see Fig. 2).

3.1 Individual Readiness Factors

There is a necessity for professionals engaged in the healthcare establishments to possess the skill in utilizing the latest technological system. There are numerous impediments for achieving it such as the know-how in terms of skills, and training of the latest systems, individual level attributes pertaining to professionals' attitudes, in addition to the manner of engaging in a task [28]. The assessment of e-Health readiness provides heightened insights regarding the understanding of readiness based on individual factors due to the significance of the association between the utilization of technology and complementing the environmental and core readiness against that of singular personal objectives [29]. For the evaluation of e-Health, the current study emphasizes an individual's features comprising: Training, Attitudes Towards and Skills.

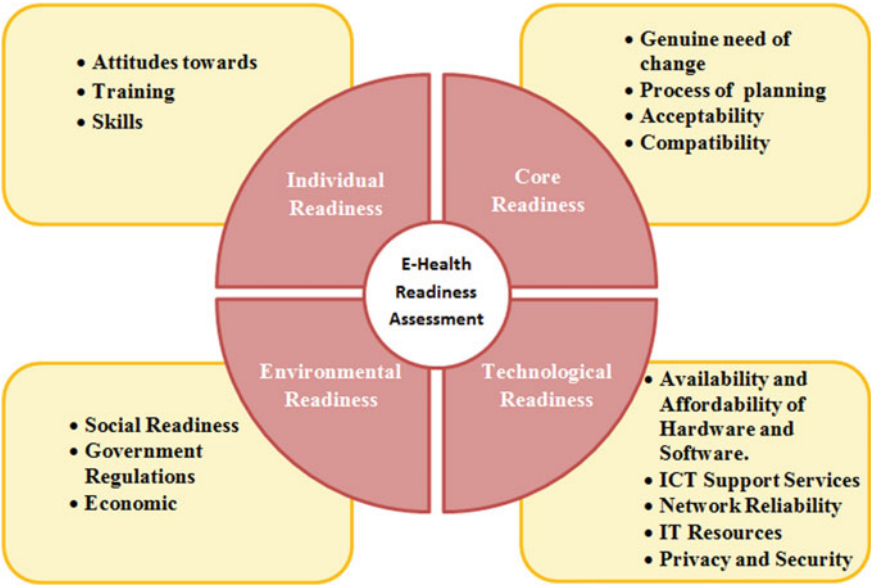


Fig. 2 Assessment of four e-health readiness

3.1.1 Attitudes Towards

The comprehension of the medical employee's attitudes in adapting with the e-Health systems is crucial in overcoming barriers as well as gain wider acceptance pertaining to the new technology employed within the healthcare sector. Furthermore, [30] asserted that a medical employee possessing the most welcoming attitude towards enhancement consequently result in a favorable impact on the adoption of institutional ground breaking inventions.

Evidentially, a positive correlation is established by such most welcoming attitude towards the acceptance of institutional innovation alongside the adoption of Electronic Health systems according to [31]. Thus, the assurance of the inclusion of users (medical care experts) associated with the e-Health system is crucial during the preceding execution stage, so as to identify their resistance and concerns early, and to rectify the same before adopting the new e-Health system. By equipping the users to be ready for the e-Health system enhances the chances of the undertaking success considerably.

3.1.2 Training

The provision of seminars and workshops for decision makers and medical personnel is very crucial, since valuing and using the technology benefits in the field of medicine causes the efficient use of e-Health and systems facilitation [32]. There is

a growing acknowledgement that training proves to be crucial for transition to the successful e-Health application [33]. Good quality training can assist tremendously in decreasing the employees' anxieties regarding the use of an unfamiliar system. The accessibility to training and technical support at the time of the initial execution of the system is highly essential. The technology presents the necessary tools in making over the routine activities, and training provides the personnel with the necessary skills. Training should be short, specialized, time-specific, and particularly designed for the specific needs of the personnel. High-quality training offers more than just skill-building; it prolongs the opportunities for communication and participation. Several surveys discovered that the absence of training was a substantial obstacle in the use of e-Health applications [34]. The current study's objective is to gain the comprehension on the issues of the users related to training in the individual readiness context.

3.1.3 Skills

The factors like the shortage of computer-literate physicians as well as medical personnel, old hardware, in addition to the lack of software use and know-how of e-Health applications have delayed the effectual employment of the new Electronic Health applications. There is a requirement for the e-Health system to be user-friendly for the doctors and the other medical personnel to make it successful [19]. The issues that lead to poor system implementation involve both the fundamental and the complex technology that the doctors and the medical personnel are incapable of managing, partially because of their lack of computer proficiency. Nonetheless, it is a fact that the majority of the providers in the healthcare domain do not have computer skills which are vital in managing the electronic instruments connected with the e-Health systems. Since certain physicians are elderly, they do not prefer to take any training in the innovative tools even though they are highly proficient in their own field. In centers where the e-Health applications were being used, issues such as lack of proficiency to manage the new technology were widespread among the medical practitioners [35].

3.2 Core Readiness Factors

Core readiness means realizing the 'genuine need' for the evaluation of Electronic Health applications, in addition to gauging the discontent pertaining the current scenario plus situations. It was ascertained in earlier researches in various fields [36]. In general, the need was seen as a basic part of readiness of a group of people to make use of e-Health: "the more the need, the more the readiness" [20]. Dissatisfaction with the present state of affairs and the readiness to use e-Health are considered fundamental to an efficacious execution of Electronic Health.

It should be noted that, certain findings reveal that the needs of the healthcare facilities are not adequate by themselves to develop willingness for the execution of e-Health [37]. The core readiness evaluation comprises four crucial factors, namely: Genuine need of change, Process of planning, Acceptability and Compatibility.

3.2.1 Genuine Need of Change

The preparedness for transformation is crucial, in addition to its being a primary phase of the effective implementation and acceptance of this innovative technology [38]. The 'Genuine Need of Change' was clear to be the model that is applicable in the willingness assessment in medical facilities when the results of the study discovered that there is an extensive scarcity of physicians in these centers and, due to this, the healthcare delivery in these centers is adversely affected. E-Health applications are viewed as an intermediary between the scarcity of doctor, and quality health care delivery [39]. Those employees who are less qualified than the physicians can carry out healthcare activities when doctors are not present by using e-Health [40]. The alterations in the current business procedures and provisions of appropriate training, in addition to the assistance for the employment of the latest practices are incorporated in the course of adoption [12]. It enables the medical centers to define the population of their clients along with their health issues, pertaining how they can be solved, and the reason of e-Health being the best to fulfil these needs [38].

3.2.2 Process of Planning

The past studies point out that with the aim to effectively adopt the innovative technology and incorporate them with the current ones, medical centers must be ready for change and include it in their planning [41]. Thus, transformation in these centers is usually a constant and progressive procedure instead of being a single occurrence, emphasizing the continuous readiness for change [14]. These notions point out that handling the aspects that determine the availability of e-Health in developing nations can enable personal and organizational willingness in medical centers, which subsequently can facilitate the growth of these facilities through various change phases. This strategy should involve evaluation and assessment of needs.

Organizational e-Health guidelines and processes must be enhanced using the strategic plan and must provide support to the overall direction of the organization. It was identified that an initial approach for strategic plan implementation, with certain milestones and reviews to be a critical part of willingness of planning [42]. Earlier researches pertaining to this planning procedure concurred with the problems recognized under the 'process of planning' section, specifically the elements associated with designating an individual or a group of individuals to assume the responsibility for plans of e-Health services and including each of the stakeholders in the process of planning.

3.2.3 Acceptability

The degree of acceptance by patients, personnel or other staff who are confident of services rendered, or that they are ready to utilize the new Electronic Health application entails 'Acceptability' [43]. Thus, the acceptance of the medical personnel towards e-Health application impacts on its success [44]. Acceptability includes perceptions of satisfaction, experience, significance and cultural suitability. The viewpoints of the personnel are really significant aspects to be considered in the assessment of the new technology due to their impact on the employment, in addition to the acceptance of the e-Health applications. The technology suitability indicates the opinion regarding the use of the new technology in certain situations [21]. Thus, the system acceptability to evaluate the e-Health readiness is utilized in this investigation.

3.2.4 Compatibility

It indicates "the extent to which organizational execution of the new e-Health applications is viewed as being coherent along with the use of the current systems, processes, and the present requirements and goals of the organizations" [45]. It has been regarded as a crucial aspect for several of the organizations' adoption of the new system. The cause of the unsuccessful adoption was the absence of consistency between the current processes and the new technology [44]. This absence of consistency could be because of various reasons. The fitting of the new e-Health system with current practices and ideas as well as previous experiences or with previously believed notions of the institutions can also enable or hinder its implementation. The study indicates that the previous experiences of the potential users with related technologies can cause both favorable and unfavorable results [46, 47]. Therefore, it is crucial for the healthcare organizations that the new system is compatible with their present needs and processes.

3.3 Technological Readiness

Technological readiness represents the capability of existing medical organization software and hardware in facilitating innovation in the clinical information technology. Thus, evaluating the existing technology readiness in the institution is of highest importance to guarantee its eventual adaptability [48]. It is possible to segregate and break down Technological readiness further into significant technological readiness concepts which are needed to select the ultimate efficient technology for the present mission, as well as for the enabling of effective adoption of the e-Health applications [49]. The respective essential technological readiness variables are: Availability and affordability of hardware and software, ICT support service, Network reliability, IT resources, and Privacy and security.

3.3.1 Availability and Affordability of Hardware and Software

The IT tools for managers has included this variable, where they handled issues of affordability, in addition to software availability of the software and hardware needed to adopt the new e-Health systems. In particular, the technological readiness variables enable the medical personnel access to new technology together with parameters of availability like the enhancement of capacity, accessibility, and suitable devices [50]. It is imperative for the devices to be suitable for task usage, positioned in handy spots, in addition with no safety hazard to utilize. The utilization of device resulted in a couple of other issues towards the favorable e-Health adoption as well as use: importance of using apt technology during instances of the provision of e-Health services are provided, and facilitating creative utilization of equipment If the equipment is very complex for the users, there will be poor use of the services, or they will not be used at all. It is crucial that medical organizations ensure that the equipment is faultless and offer technical support to the users. In isolated regions, it is critical to make schemes for the upkeeping of device, in addition to resolution of technological problems [51].

3.3.2 ICT Support Service

ICT support represents the provision of various technical support facilities to the users by the organization. Technical support is regarded as a critical requirement for both IT managers and medical staff in order to utilize the new e-Health application efficiently [52]. Moreover, medical organizations possessing extensive IT support would undergo an easier changeover to the e-Health system. Medical organizations that obtain the software technology from huge experienced developers stand a better chance of software adoption in initial stages itself because of the substantial IT support possessed. Bigger medical organizations have more resources available to them, and they also possess more practical experience for enabling the institutions to make transition to software system. Smaller medical organizations are short of much support as well as resources required to efficiently help organizations to customize their use of the e-Health application [53]. Thus, this study considers this aspect as a necessary construct influencing the present e-Health readiness in the medical establishments of Iraq to use technology.

3.3.3 Network Reliability

In this research, the telecommunication infrastructures function as the network that connects the various healthcare sectors and users in a region or within a state. As stated by [54] no Internet connection is one of the principle impediments that restrict the new technologies acceptance and utilization. A number of other researchers also mentioned that there are certain challenges that the developing countries face with regards to Internet accessibility. The problems or issues are due

to inadequate telecommunication capabilities and erratic power supply. However, as mentioned by [37], access to a reliable network in addition to the exchange and dissemination of medical data using that network is presently hindered by the continuous manual use of printed materials which are preserved and managed in the majority of the medical information. Moreover, according to the observations made by [16], medical data remain to be maintained and kept in incongruent body of data record in diverse formats, although in certain instances computerization were enabled. This leads to considerable difficulty in offline mode of working, for the medical officials working in regions with absence of network connectivity. Hence, this factor is of significant relevance and impact with respect to the use of e-Health services against the background of developing countries.

3.3.4 IT Resources

The aim of the current research is in ascertaining the components which influence the technological readiness, the degree of preparedness for the advancement of wellbeing amenities may be affected. It was observed that IT resources play a significant role when developing technological readiness [55]. Based on earlier research, the assessment of IT resources has been conducted, which is established to be a crucial element in the evaluation of Electronic Health implementation readiness [9, 22]. IT resource readiness applies to the degree of infrastructure capability that can support the implementation of a clinical IT/S innovativeness. The assessment can be carried out by decision-makers who are aware and well-versed with the types of organizational resources that are available for the initial adoption of IT/S innovation and its eventual implementation, as well as for the on-going maintenance of the innovation. It is necessary to have a definite level of IT resource readiness to ensure that healthcare providers are comfortable and confident to be engaged in the employment of Electronic Health services [19]. In certain studies, related to the wellbeing industries, the significance of IT resources was highlighted to be an essential element in the assessment of Electronic Health preparedness, in being a major component of the development plans.

3.3.5 Privacy and Security

In order to develop the confidence amongst the medical staff and the patients, the security of data alongside privacy are two factors that play an imperative role. The specification of privacy is the entitlement in addition to the aspiration of a person to manage the revelation of personal health data. On the other hand, information security is defined as an aggregate of guidelines, processes, as well as safeguard measures which assist to conserve the integrity of, and regulate admission and availability of their contents. Both the parameters of privacy and information security require fortification of legislative structure that promotes the safeguarding

of privacy, in addition to the development of efficient attestation government bodies, and authorization of digitized signatures, which will support the fortification of the trust and security infrastructure [56].

Even though the accessibility of the contents of patient PHR (personal health records) are needed to be available to varying wellbeing care providers and specialists, the extent and degree of accessibility need to be regulated. Most of the research on the preparedness to employ Electronic Health systems has validated the said outcomes [13, 57]. Hence, it is crucial for any e-Health readiness assessment to possess established procedures and standards for the safeguard of privacy as well as data at the time of the installation of an e-Health program has been set up for the routine delivery of healthcare.

3.4 Environmental Readiness

The external variables which are present apart from the scope of control of the management of the hospital or the employees of the hospital are referred to as the environmental factors. The core objective of the environment is to allow the IT services to meet the present necessities and needs, and to ensure that there is no impact or compromise on the ability of future generations to address there. The various concerns associated with wellbeing establishment include Government Regulations, Social Readiness, and Economic Readiness.

3.4.1 Social Readiness

Social or community readiness is signified to be the extent healthcare establishment, as well as its employee willingness to contribute and engage in the interconnected online realm that involves the collaboration in addition to the communication amongst the center of administration and the hospitals, and the establishment of healthcare internally and external of the country [58]. The different attributes that are necessary for healthcare institutions include sharing of information, collaboration with other healthcare institutions, and provision of patient care alongside the societies in cooperation with different establishments. In addition, Societal Readiness takes into account societal and cultural concerns amongst the employee and clients. According to certain studies by [50], the societal norms have a considerable influence on the adoption of IT practices, particularly in Arab countries. To successfully build and use an information system of significance, the change leader is required to assess how that information system will impact the work performed within the existing system by the individuals and the groups to which they belong. Hence, the assessment the preparedness of the e-Health system through the evaluation of societal readiness is crucial so that the group values of which the information system is required to support can be identified.

3.4.2 Government Regulations

Government regulations refer to a collection of directives, statements, legislations by the government in addition to legal interpretations which command and handle the e-Health system course of life [10]. The said rules may be a restraining variable entrenched in a specific sector that is not open to the utilization of new innovations; meanwhile in a different sector, these rules and policies could demonstrate to be encouraging for organizations to utilize innovative technologies [59]. It implicates on the readiness of the nation, with respect to the government's commitment in promoting, supporting, facilitating and regulating e-Health with its numerous needs, in addition to providing it with a legal infrastructure, is of considerable significance. Governmental entity plays a key role in the decision-making process of implementing a novel scheme in healthcare establishments. The use or adaptation of treatment or managerial tools is governed by regulations that are increasingly pervaded by aspects connected with transformations in technology [60].

The government interventions and regulations develop incentives and constraints that influence the process of technological change and adoption. In various studies, this factor has been analyzed and the findings related to its importance and part played in terms of e-Health readiness evaluation were not in agreement due to the variations in the environmental factors and settings [13].

3.4.3 Economic

The economic factor refers to the extent of the policy makers' perception towards the overall expenditure of the implementation and utilization of e-Health system being less than the use of alternative traditional systems [61]. The economic factor takes into consideration the overall cost that includes fixed costs such as initial investment, training costs, operating costs, cost of migration from the old systems to the new, cost of system maintenance and upgrades, and variable costs. For example, the minimal cost of software and hardware that is selected also provides advantages such as reduced power utilization that is kept to a minimal, and assurance that the employment of resource remains at high levels [10]. The economic factor also considers the replacement costs and costs of managing the life cycle [62]. Cost of care refers to the worth of resource utilized economically which is linked to the objective of defining findings, it measures the worth of resources consumed due to a goal or initiative. In this age where budget constraints are a major issue, the costs and efficient expenditure related to the Electronic Health system in relation to the traditional health services are justifiably significant for decision-makers. Hence, the economic factor measures the value of resource use related to an intervention.

4 The Interview

The interview was conducted with the objective of obtaining thorough information about the dimensions of the research problem, from the target population. In the present study, the semi-structured interview approach was used by the researchers to ask more questions on the predefined questionnaire. The purpose of the interview was also to ascertain whether there were any latent factors present that influence the e-Health readiness assessment in healthcare institutions in Iraq. In addition, interview was aimed at compiling the opinion of the IT specialists in the healthcare institutions on the IT services and environments, and the constraints of IT departments. The interview was also aimed at analyzing the advantages and shortcomings of the existing IT system from the institutions that have implemented other IT architectures. The factors that prove to be detrimental in supporting the assessment of e-Health readiness are also identified in this study. Therefore, the interview was conducted with IT managers and clinical staff for the attainment of the principal goals of this research.

The researcher directed questions in a semi-structured form to a group of eight experts that included medical personnel in addition to technical personnel. The medical and technical staff who were interviewed were from health care institution in Iraq, situated in the Baghdad Governorate. The primary rationale for this selection is due to the fact that the technology adapted in this region was superior in terms of functionality and reliability, compared to that in the other states of Iraq. Approximately, there are about 39 hospitals in Baghdad. However, for this study, members of the staff were selected from only six hospitals. The choices were based on fact that this study required medical professionals and technical experts who were familiar with the use of technology.

The interview questionnaire focused on the concerns of the hospitals in Iraq associated to the utilization of e-Health services. The questions were developed on the basis of the highlights specified by the hospital managers. They helped the researcher to recognize the key concerns related to the organizational and individual dimensions. There were two additional experts from the same hospitals who validated the factors identified based on the managers' inputs. This process was essential to ensure the questions were consistent, as well as in accordance with the objectives of the current study. Appendix A provides the details of the questions.

5 Conclusion

This study was carried out to evaluate the Electronic Health readiness so that the potential causes of failure in relation to the implementation of technological innovations in the Iraqi healthcare sector could be identified. It is evident from the literature that the implementation of technological innovation begins with the assessment for e-Health readiness which enables and prepares the organizations to get associated with the networked community. The findings of the preliminary study are based on a semi-structured interview involving eight medical personnel

and technical personnel from six hospitals in Iraq. The resultant factors were in accordance with the conclusions made by the literature, in which the mentioned factors were found to be most relevant in connection with the current research.

This study analyses key variables that impact the environmental, technological, individual, and core readiness, and subsequently impact the utilization of e-Health in the healthcare industry in Iraq. It can be concluded from the findings that a majority of the interviewees agreed with these factors which pose a substantial impact on the present state of utilization of e-Health systems. The current study also analyzes the present situation of using e-Health services in the public healthcare industry in Iraq. Moreover, the observations made in this study can contribute towards achieving a high utilization of other e-health initiatives in Iraqi healthcare industry such as telemedicine, e-Health records, as well as the implementation of cloud computing.

Appendix A

This section presents the interview questions used to interview the participants.

Q1: How the treatment and health process of healthcare providers, particularly within Iraqi healthcare institutions, be enhanced in the 21st century?
Q2: Enumerate the complexities and difficulties associated with implementing and understanding a new information system within healthcare organizations?
Q3: What strategic planning issues need to be considered by healthcare institutions when it comes to implement a new information system within these institutions?
Q4: What individual qualities do you think are vital in influencing the assessment of e-Health readiness to use a new information system?
Q5: What important essence technological issues strategic to the e-Health system readiness are vital for assessment consideration?
Q6: What core issues do you consider important regarding e-Health system and how it is by healthcare institutions?
Q7: What environmental and external factors need to be considered regarding the assessment of e-Health readiness in healthcare institutions?
Q8: What other information such as issues, ideas, or recommendations would benefit this study?

References

1. Zonneveld M et al (2020) The use of information and communication technology in healthcare to improve participation in everyday life: a scoping review. Disabil Rehabil 42 (23):3416–3423
2. Kadhum AM, Hasan MK (2017) Assessing the determinants of cloud computing services for utilizing health information systems: a case study. Int J Adv Sci Eng Inf Technol 7(2):503–510

3. Noor A (2019) The utilization of e-health in the Kingdom of Saudi Arabia. Int Res J Eng Technol 6:1229–1239
4. Beebeejaun MR, Chittoo H (2017) An assessment of e-health readiness in the public health sector of Mauritius. Int J Sci Basic Appl Res 35(1):193–210
5. Mauco KL, Scott RE, Mars M (2020) Validation of an e-health readiness assessment framework for developing countries. BMC Health Serv Res 20(1):1–10
6. Al Hilfi TK, Lafta R, Burnham G (2013) Health services in Iraq. The Lancet 381(9870):939–948
7. Hameed IH et al (2015) Alkaloid constitution of *Nerium oleander* using gas chromatography-mass spectroscopy (GC-MS). J Med Plants Res 9(9):326–334
8. Kutlu B, Ozturan M (2012) Determinants of E-Health readiness of end-users. Int Med J 19 (4):287–291
9. Durrani H et al (2012) Health needs and eHealth readiness assessment of health care organizations in Kabul and Bamyan, Afghanistan/Évaluation des besoins en matière de santé et de la préparation à la cybersanté dans des établissements de soins de santé à Kaboul et Bamyan (Afghanistan). East Mediterr Health J 18(6):663
10. Rezai-Rad M, Vaezi R, Nattagh F (2012) E-health readiness assessment framework in iran. Iran J Pub Health 41(10):43
11. Li J et al (2012) Issues regarding the implementation of ehealth: preparing for future influenza pandemics. Interact J Med Res 1(2)
12. Khoja S et al (2013) Conceptual framework for development of comprehensive e-health evaluation tool. Telemed e-Health 19(1):48–53
13. Ghazisaeidi M et al (2014) An assessment of readiness for pre-implementation of electronic health record in Iran: a practical approach to implementation in general and teaching hospitals. Acta Med Iran 52(7):532
14. Qureshi QA et al (2013) Determining the users'willingness to adopt electronic health records (ehr) in developing countrIES. Gomal Univ J Res 28(2):140–148
15. Buabbas A (2013) Investigation of the adoption of telemedicine technology in the Kuwaiti health system: Strategy and policy of implementation for overseas referral patients. Brunel University, School of Information Systems, Computing and Mathematics
16. Coleman A, Coleman MF (2013) Activity theory framework: a basis for e-health readiness assessment in health institutions. J Commun 4(2):95–100
17. Kgasi M, Kalema B (2014) Assessment E-health readiness for rural South African areas. J Ind Intell Inf 2(2)
18. Abd Ghani MK, Jaber MM (2015) Willingness to adopt telemedicine in major Iraqi hospitals: a pilot study. Int J Telemed Appl 2015
19. Alghamdi AS (2015) Factors associated with the implementation and adoption of electronic health records (EHRs) in Saudi Arabia. Rutgers University-School of Health Professions
20. Almuayqil S, Atkins AS, Sharp B (2016) Ranking of e-health barriers faced by saudi arabian citizens, healthcare professionals and IT specialists in Saudi Arabia. Health 8(10):1004
21. Mauco KL, Scott RE, Mars M (2016) Critical analysis of e-health readiness assessment frameworks: suitability for application in developing countries. J Telemed Telecare 1357633X16686548
22. Ahmadzada S, Zayyad MA, Toycan M (2016) Readiness assessment for the use of cloud computing in eHealth systems: a field study of hospitals in the capital of Azerbaijan. In: HONET-ICT, 2016. IEEE
23. Meri A et al (2019) Modelling the utilization of cloud health information systems in the Iraqi public healthcare sector. Telematics Inform 36:132–146
24. Ritchie J et al (2013) Qualitative research practice: a guide for social science students and researchers. Sage Publications
25. Yin RK (2013) Case study research: design and methods. Sage Publications
26. Jacob SA, Furgerson SP (2012) Writing interview protocols and conducting interviews: tips for students new to the field of qualitative research. Qual Rep 17(42):1–10

27. Braun V, Clarke V (2006) Using thematic analysis in psychology. Qual Res Psychol 3(2):77–101
28. Yusif S et al (2020) PLS-SEM path analysis to determine the predictive relevance of e-Health readiness assessment model. Health Technol 10(6):1497–1513
29. Adeola O, Evans O (2018) Digital health: ICT and health in Africa. Aktual'ni Problemy Ekonomiky= Actual Prob Econ 208:66–83
30. Muvoti M (2018) *An analysis of the diffusion and adoption of technologies and innovations in the public healthcare sector in Zimbabwe*
31. Hossain N et al (2019) Factors influencing rural end-users' acceptance of e-health in developing countries: a study on portable health clinic in Bangladesh. Telemed e-Health 25 (3):221–229
32. Musen MA, Middleton B, Greenes RA (2014) Clinical decision-support systems. Biomedical informatics. Springer, pp 643–674
33. Riddell MC et al (2014) Achieving meaningful use of electronic health records (EHRs) in primary care: proposed critical processes from the Kentucky Ambulatory Network (KAN). J Am Board Family Med 27(6):772–779
34. Sardi L, Idri A, Fernández-Alemán J (2017) A systematic review of gamification in e-Health. J Biomed Inform 71:31–48
35. Paré G et al (2014) Barriers to organizational adoption of EMR systems in family physician practices: a mixed-methods study in Canada. Int J Med Informatics 83(8):548–558
36. Davis KM (2016) Using the transtheoretical model of change to improve lifestyle behaviors in a woman with metabolic syndrome. J Doctoral Nursing Pract 9(1):81
37. Justice EO (2012) E-healthcare/telemedicine readiness assessment of some selected states in Western Nigeria. Int J Eng Technol 2(2):195–201
38. Gagnon M-P et al (2014) Electronic health record acceptance by physicians: testing an integrated theoretical model. J Biomed Inform 48:17–27
39. Akacho EN (2014) Factors influencing provision of health care service delivery in Kenya. a case of Uasin Gishu District hospital in Eldoret. Unpublished research project, University of Nairobi
40. Soltysik A (2016) Organization of healthcare entity: the study of hospital information system. In: *R.* Innowacje w zarządzaniu i inżynierii produkcji, PTZP, Opole, pp 141–147
41. Eldredge LKB et al (2016) Planning health promotion programs: an intervention mapping approach. Wiley
42. Scharwz F, Ward J, Willcock S (2014) E-Health readiness in outback communities: an exploratory study. Rural Remote Health 14(3):2871
43. Cameron P, Munyan K (2021) Systematic review of telehospice telemedicine and e-health. Telemed e-Health
44. Dixon BE, Vreeman DJ, Grannis SJ (2014) The long road to semantic interoperability in support of public health: experiences from two states. J Biomed Inform 49:3–8
45. Rivas R, Vilcahuamán L (2017) Healthcare technology management systems: towards a new organizational model for health services. Academic Press
46. Harris ML, Gibson SG, Taylor SR (2015) Examining the impact of small business institute participation on entrepreneurial attitudes. J Small Bus Strateg 18(2):57–76
47. Liedtka J (2015) Perspective: linking design thinking with innovation outcomes through cognitive bias reduction. J Prod Innov Manag 32(6):925–938
48. Chen Y et al (2017) EHR: a sensing technology readiness model for lifestyle changes. Mob Netw Appl 1–15
49. Wouters EJ, Weijers TC, Finch TL (2017) Successful implementation of technological innovations in health care organizations. In: Handbook of smart homes, health care and well-being, pp 179–189
50. Alaboudi A et al (2016) Barriers and challenges in adopting Saudi telemedicine network: the perceptions of decision makers of healthcare facilities in Saudi Arabia. J Infect Public Health 9(6):725–733
51. Busch J (2014) The impact of ICT enabled process innovation on efficiency measures

52. Taylor J et al (2015) Examining the use of telehealth in community nursing: identifying the factors affecting frontline staff acceptance and telehealth adoption. J Adv Nurs 71(2):326–337
53. Hoffman S, Podgurski A (2012) Balancing privacy, autonomy, and scientific needs in electronic health records research
54. Meri A, Hasan M, Safie N (2018) The impact of organizational structure and system settings on the healthcare individual's perception to utilize cloud services: a theoretical literature survey. J Eng Appl Sci 13(4):888–897
55. Abd Ghani M, Neamah AF (2016) Electronic health records challenges and barriers in Iraq. Comput Eng Intell Syst 7:2222–2863
56. Soceanu A et al (2015) Managing the privacy and security of ehealth data. In: 20th International conference on control systems and computer science (CSCS). IEEE
57. Pankomera R, van Greunen D (2016) Privacy and security issues for a patient-centric approach in public healthcare in a resource constrained setting. In: IST-Africa Week Conference. IEEE
58. Ali A (2019) e-Health systems adoption and telemedicine readiness: practitioner perspective from Libyan healthcare sector. Brunel University, London
59. Kim G-H, Trimi S, Chung J-H (2014) Big-data applications in the government sector. Commun ACM 57(3):78–85
60. Kraft MME (2017) Environmental policy and politics. Taylor & Francis
61. Boore C (2018) A framework for e-health implementation in the health care sector in Kenya: a grounded theory approach. University of Nairobi
62. Brata C et al (2013) A review of the information-gathering process for the provision of medicines for self-medication via community pharmacies in developing countries. Res Social Adm Pharm 9(4):370–383

Toward Iraqi Healthcare System Framework

Saif Mohammed Ali, Sura Khalil Abd, Mustafa Musa Jaber, and Ali Taha Yaseen

Abstract E-Health is an application of Information and Communication Technologies (ICT) across the range of functions which affect health for example, treat patients, pursue research, educate students, track diseases and monitor public health. E-Health has become a sturdy and versatile platform that can facilitate the enhancement of information and services delivery in the health industry. E-Health readiness assessment is a basis for ICT investments in healthcare organizations. In relation to healthcare, assessing health readiness is impacted by several factors. However, even though there is an abundance of evidence of the subject of healthcare in the literature, studies that have investigated the factors impacting its assessment in the developing countries is still limited. In Iraq, e-Health readiness is still in its nascent stage. Therefore, this study aims to assess the contribution of the critical factors leading to e-Health readiness assessment in healthcare institutions in Iraq. The propose study also contributes to the body of knowledge among medical and technical staff with regard to their beliefs in the use of e-Health. In addition, these significant findings have crucial implications and valuable contributions to researchers, government agencies, insurance companies, and healthcare professionals in Iraq.

Keywords First keyword · Second keyword · Third keyword

S. M. Ali (✉) · M. M. Jaber · A. T. Yaseen
Computer Science Department, Dijlah University College, Baghdad, Iraq
e-mail: saif.alameri@duc.edu.iq

A. T. Yaseen
e-mail: ali.taha@duc.edu.iq

S. K. Abd
Computer Technologies Engineering Department, Dijlah University College, Baghdad, Iraq
e-mail: Sura.khalil@duc.edu.iq

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_57

1 Introduction

Generally, e-Health systems are utilized in the healthcare sectors for the purpose of supporting research activities, teaching, and public services. The efficiency of computing application in health sectors enables healthcare professionals to control and manage patients' health records [1]. Electronic health systems can be effectively collected, stored, processed and exchanged using Information and Communication Technologies (ICTs). It has been shown that healthcare services can be improved in quality, safety and with reduced cost using ICT. The implementation of technology in solving healthcare-related problems opens up commercial opportunities and resided efficiency level of healthcare delivery across both developed and developing nations.

Huge advances have been made in information technology especially in the health sector in the last two decades, with several designs of electronic health devices and their implementation [2]. Since the need for healthcare services are continuously increasing, it is important for the provision of healthcare services to be efficient and readily accessible [3]. Healthcare based on the ICT (as e-Health) is promising and may help to overcome problems associated with the conventional (paper-based) healthcare. By definition, e-Health is the application of ICT in the process of delivering healthcare, such as in the treatment of patients, research purposes, disease monitoring, training, and public health monitoring. More research attention is now given to e-Health (or e-Healthcare) owing to its broad patient involvement.

2 Background of e-healthcare System

Organizations, professionals and especially patients and the public are expected to benefit from the implementation of e-Health. The review of the literature confirms the many significant benefits of e-Health to patients [4, 5]. A major benefit that has been reported is the improvement in quality of care as a result of the ease of accessibility to patients' essential health data by their different health providers. Efficient, appropriate and the latest information can be provided to healthcare professionals for knowledge exchange and decision-making using e-Health systems. Also, clinical practices, distribution of health services, quality of care, duplication of medical examinations, wrongful dispensation of medicines and patient safety can all be improved with e-Health systems. Moreover, patient safety can be protected by implementing the right e-Health system in the right way. The assessment of e-Health readiness will allow organizations facilitate successful implementation. E-Health readiness is defined as the readiness of healthcare institutions and communities to receive the expected changes occurring from the introduction of ICT programs [6].

Assessment of e-Health readiness should serve as part of pre-implementation evaluation to ensure that e-Health systems are successfully implemented. For example, possible reasons for not innovating can be identified from e-Health readiness assessment. Evidence to support the readiness (or lack of it) of an organization to be transformed before implementing e-Health, can be determined from e-Health readiness assessment [7].

3 Related Works

The background of the study provides information concerning the challenges in implementation of the e-Health system. E-Health for healthcare has evolved due to the revolutionary advances made in ICT network technologies and the improvements in information management. The new technologies have transformed healthcare delivery and modern technology has the capability to support and manage the various aspects of healthcare, ranging from searching general health information to consultations with physician without the need for patients to leave their homes. However, some limitations still exist, one of which is the lack of accessibility of patients to their most basic personal health information, such as test results, which are present in medical charts but unavailable to patients via e-Health technologies [8]. Thus, patients become inactive participants in e-Healthcare delivery due to the lack of access to their own information.

Readiness, which is carried out in the early stage of implementation of change is the cognitive precursor for measuring the support-for or resistance-to change [9]. This study was focused on readiness involving health workers' readiness and organizational readiness. Assessment of health workers' readiness relates to the e-Health record practices in hospitals (such as knowledge in, access to or difficulties in using computers) while organizational readiness involves the evaluation of organizational resources (ICT infrastructure and finance required to implement e-Health).

Readiness assessment is required before implementing e-Health systems in healthcare organizations. Risk of failures in organizational projects can be reduced using readiness assessment. It is important to check the availability and local context of basic factors that will promote acceptance and utilization of e-Health. User information, organizational strengths and weaknesses can all be identified from readiness assessment [10]. E-Health systems are capital intensive and their failure can be devastating for the implementing organization, which justifies the importance of readiness assessments (technical and social) prior to actual implementation.

Nonetheless, the recommendation of readiness assessment is often neglected in the implementation of ICT projects in public health sector for political as well as socio-cultural reasons [11]. Developing countries often lack some basic infrastructures and policies needed for implementing health-related projects. It therefore is highly important for healthcare institutions in developing countries preparing to

adopt e-Health to conduct readiness assessment studies, in order to evaluate the organizational and local factors that are available or may be needed to facilitate the acceptance, utilization and sustainability of such initiative [12].

These background studies have presented empirical confirmation from research carried out in the assessment of e-Health readiness. The findings from these previous researches done will offer information on contemporary trends in communication and systems implementation on national and global levels.

4 Significance of the Study

This study will provides sufficient understanding about the state of the current healthcare structure and system characteristics necessary for accommodating individual needs to be willing to assess and improve the readiness of e-Health in Iraqi healthcare organizations. Healthcare professionals have the responsibility to ensure consumers are knowledgeable in respect of the changes implemented for the use of e-Health, access to personal health information, and consumer rights for privacy and security.

The data will collect from this study would be beneficial for government agencies and e-Health stakeholders. The government has taken advantage of technology by providing information on policy and services offered by governmental agencies, such as public health agencies, ministries of health, healthcare providers, international organizations, donor countries, aid agencies, which have taken on new roles to provide leadership in terms of strategic direction as it affects the entire nation and its healthcare sector [13]. Moreover, it also could be beneficial for researchers, health planners, academics, students, and other e-Health stakeholders who may need to know a country's e-Health readiness level. Having available an evidence-based and practical tool to assess e-Health readiness of a country would offer a valuable avenue in assessing the feasibility of e-Health implementation in a country and thereby facilitate investment in, and growth and global integration of e-Health.

5 The Aim of This Research

In light of the research problems and purpose mentioned above, the research questions are formulated to in order to attain the research objectives. First is to investigate the current state and level of e-Health system among the healthcare providers in Iraq. Then discover the critical factors for the readiness assessment of e-Health in Iraqi healthcare organizations. Finally propose a model for assessing the e-Health readiness in the Iraqi healthcare sector. To validate the proposed model for e-Health readiness assessment in Iraqi healthcare sector.

6 Issues Facing e-Health System

An e-Health system refers to "the utilization of IT-related systems in order to capture, store, manage or transmit information related to the health of individuals or the activities within the health sector" [14]. For the time being, e-Health systems utilize the delivery of quality healthcare to the people as well as address numerous problems related to healthcare service and health information system. In evolving from a paper-based system to an e-system, and determining the most suitable technological system is a daunting task. Therefore, healthcare organizations are attempting to identify the latest and the best technological innovation in their strategic investments to for greater competitiveness [15].

The use of e-Health systems by healthcare professionals is still limited, despite the urgent need to find a lasting solution to issues of medical error from consumers and regulatory agencies. It is estimated that 100,000 deaths occur due to medical errors (The Pioneer Institute [16]). In developing countries, medical technology-based projects are often implemented without sufficient assessment of local needs or comparison with alternative forms of treatment. The fact remains that "old" technologies are being used to only a fraction of their potential. Proposed projects in developing countries have had a history of overlooking the urgent needs of a countrywide equipment maintenance service, often resulting in closets filled with equipment that is useless or unusable. The cases of failure in health technology projects are well documented, but a recent publication nevertheless maintains there is the potential for technology to redefine readiness assessment in the health sector in developing countries [17].

Since e-Health systems have many, healthcare institutions in developing countries with insufficient financial resources for IT infrastructures and services, will gain more from its use [18]. As these healthcare organizations need to strategies in order to use their investment profits to achieve a competitive advantage. The literature has admitted to the challenges in deciding to adopt e-Health systems without assessing the readiness of changing the existing e-Health systems [5, 19].

Many studies recommended a full utilization of e-Health systems in order to improve the efficiency and performance of Healthcare services in Iraq [20, 21]. However, despite the many benefits that can be obtained from the adoption of ICTs in the healthcare sector, the implementation process without understanding the critical factors related to assessment of e-Health readiness is frequently resisted and has proven to be extremely difficult which indicates that the successful implementation of these projects is not a straightforward or an easy task for Iraqi healthcare organizations [22].

7 Issues Facing e-health in Iraq

Reference [14] maintained that the existing healthcare sector in Iraq continued to face many issues to improve the e-Health system and assess the health professionals to improve the organizational performance. They indicated the need to upgrade the quality of data as it moved up the chain from facility to Ministry of Health on the basis of the health facilities being unequally spread in meeting the country's health demands. Reference [23] recognized the existing problems that faced the Iraqi healthcare sector in providing medical management services because of the unsuitable use of technology with the lack of a system implemented for information sharing.

The issue to be addressed in this study is the absence of understanding of the key aspects which impact the e-Health readiness assessment in healthcare organizations of developing countries like Iraq. Furthermore, the extent of readiness levels is unclear, which raises another issue for e-Health readiness assessment at come up with a highly comprehensive discussion of the aspect of e-Health systems and healthcare service delivery. Thus, this preliminary study is proposed for the development of a framework for e-Health readiness assessment of healthcare institutions in Iraq, as well as possible interrelations between these factors.

8 Contribution of the Research

The scope of research typically constitutes the first step in a literature review [24]. On the other hand, this step needs to fully understand the research field and the possible effect of the findings (HOW), the users to solve the research questions and those who stand to gain from the findings (WHO), and the potential answers to the research questions (WHAT). A definition of the research scope offers the researcher the opportunity to focus on the research questions. In this current study, only six hospitals (Baghdad Teaching Hospital, Surgical Specialties Hospital, Private Nursing Home Hospital, Child Protection Teaching Hospital, Gastroenterology and Hepatology center, and Iraqi Center for Cardiac Diseases) are featured because they are the major hospitals with advanced IT facilities in Baghdad but are challenged by time constraints, resources and problems in establishing direct contact with department managers.

9 Conclusion

The study discussed a review of the literature in association with the research proposal. Concepts regarding e-Health readiness were defined, along with discussion of the benefits of innovative e-Health technologies in providing competitive

advantages to healthcare organizations. Additionally, the e-Health readiness and implementation status of medical organizations were determined in Iraq as well as other states. The review indicated that hospitals in Iraq are sluggish in the deployment of e-Health systems. It also established the issues that thwart e-Health readiness evaluations of healthcare organizations that attempt to implement e-Health services in Iraq. As well, various aspects regarding e-Health readiness assessments of healthcare institutions were examined and mutually evaluated. The results show that every factor comprises a certain set of disadvantages and advantages. Therefore, appropriate choices would depend on the individual qualities, situations, and requirements of hospital units. Lastly, the studies most associated with this research are discussed in order to bridge gaps in current understandings, towards advancing readiness assessment models that can better assist healthcare institutions in using e-Health systems.

Acknowledgements "This work is a part of a project submitted to Universiti Teknokal Malaysia Melaka."

References

1. Meri A et al (2019) Modelling the utilization of cloud health information systems in the Iraqi public healthcare sector. Telematics and Informatics 36:132–146
2. Avgerou C, Walsham G (2017) Information technology in context: studies from the perspective of developing countries. Routledge
3. Cohen L et al (2016) Healthcare information analysis and graphical display presentation system. Google patents
4. Zayyad MA, Toycan MJP (2018) Factors affecting sustainable adoption of e-health technology in developing countries: an exploratory survey of Nigerian hospitals from the perspective of healthcare professionals. Peer J 6:e4436
5. Beebeejaun MR, Chittoo H (2017) An assessment of e-health readiness in the public health sector of Mauritius. Int J Sci: Basic Appl Res 35(1):193–210
6. Ali A (2019) e-Health systems adoption and telemedicine readiness: practitioner perspective from Libyan healthcare sector. Brunel University, London
7. Yusif S, Hafeez-Baig A, Soar J (2017) e-Health readiness assessment factors and measuring tools: a systematic review. Int J Med Inform 107:56–64
8. Furusa SS, Coleman A (2018) Factors influencing e-health implementation by medical doctors in public hospitals in Zimbabwe. S Afr J Inf Manage 2018. **20**(1): p. 1–9.
9. Arnéguy E, Ohana M, Stinglhamber F (2018) Organizational justice and readiness for change: a concomitant examination of the mediating role of perceived organizational support and identification. Front Psychol 9:1172
10. Yusif S, Hafeez-Baig A, Soar J (2020) An exploratory study of the readiness of public healthcare facilities in developing countries to adopt health information technology (HIT)/ e-Health: the case of Ghana. J Healthc Inf Res 4(2):189–214
11. Touré M, Poissant L, Swaine BR (2012) Assessment of organizational readiness for e-health in a rehabilitation centre. Disabil Rehabil 34(2):167–173
12. Mauco KL et al (2018) Critical analysis of e-health readiness assessment frameworks: suitability for application in developing countries. J Telemed Telecare 24(2):110–117

13. Abd Ghani MK et al (2008) Electronic health records approaches and challenges: a comparison between Malaysia and four East Asian countries. Int J Electron Healthc 4(1):78–104
14. Kadhum AM, Hasan MK (2017) Assessing the determinants of cloud computing services for utilizing health information systems: a case study. Int J Adv Sci Eng Inf Technol 7(2):503–510
15. Laurenza E, Quintano M, Schiavone F, Vrontis D (2018) The effect of digital technologies adoption in healthcare industry: a case based analysis. Bus Process Manage J 24(5)
16. Pioneer Institute (2009) CPOE initiative: New England Healthcare Institute and Massachusetts Technology Collaborative Cambridge [policy brief: paper no. 28]. Retrieved from http://www.pioneerinstitute.org/pdf/ROE28.pdf
17. Mauco KL, Scott RE, Mars M (2016) Critical analysis of e-health readiness assessment frameworks: suitability for application in developing countries. J Telemed Telecare 1357633X16686548
18. Ahmadi H, Nilashi M, Ibrahim O (2015) Organizational decision to adopt hospital information system: an empirical investigation in the case of Malaysian public hospitals. Int J Med Inf 84(3):166–188
19. Ahmadzada S, Zayyad MA, Toycan M (2016) Readiness assessment for the use of cloud computing in eHealth systems: a field study of hospitals in the capital of Azerbaijan. In: HONET-ICT, 2016. IEEE
20. Meri A et al (2019) Modelling the utilization of cloud health information systems in the Iraqi public healthcare sector. Telematics Inform 36:132–146
21. Al-Moosawi K, Nayyef M (2017) Survey study: reality check of internet and telemedicine use in Iraqi hospitals. J Int Soc Telemed eHealth 5:7–1–6
22. Musa Jaber M et al (2014) A review of adoption of telemedicine in middle east countries: toward building Iraqi telemedicine framework. Sci Int 26(5)
23. Hameed IH et al (2015) Alkaloid constitution of *Nerium oleander* using gas chromatography-mass spectroscopy (GC-MS). J Med Plants Res 9(9):326–334
24. Booth A, Papaioannou D, Sutton A (2011) Systematic approaches to a successful literature review. Sage

Advances in Information Systems

Adoption of Cloud-Based E-Government Services: A Bibliometric Analysis and Research Agenda

Atheer Al-Haideri, Ahmad Shukri Mohd Noor, and Ismail Ahmed Al-Qasem Al-Hadi

Abstract Studies on the adoption of cloud computing in the public sector institutions, including governments remain preliminary as most of the reported works in this aspect focus more on the technical perspective of cloud computing and the potential of saving cost through its adoption. Although numerous studies have been conducted on cloud computing adoption, a bibliometric study on the adoption of cloud computing in the public sector is yet to be conducted. Therefore, this study strives to fill this gap by embarking on a comprehensive evaluation of the adoption of cloud computing and the trends in E-government research between 2009 and 2021. To achieve this aim, bibliometric performance analysis and science mapping techniques were employed in this study to analyze most of the publications on cloud computing adoption in the public institutions, focusing more on the most coted ad most productive countries, the commonly used keyword, and the most cited authors in this context. A research agenda was also presented in his study with the intent of contributing perspectives for further studies in this area.

Keywords Cloud computing · E-government · Adoption · Bibliometric analysis · Public sector

1 Introduction

The provision of on-demand computing resources, such as storage, servers, databases, servers, software, etc. over the Internet on a pay-for-use basis is referred to as cloud computing [1, 2]. Cloud computing users can control these computing resources from anywhere at any time through an interface that facilitates its usage [3]. E-government can be considered a tool that improves the way of governance and encourages the participation of the citizens in governance; it can improve transparency and efficiency in public service delivery [4, 5]. The potential

A. Al-Haideri (✉) · A. S. M. Noor · I. A. A.-Q. Al-Hadi
Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Kuala Nerus, Terengganu, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_58

opportunities provided by ICT have been recognized by most governments worldwide in the aspect of improving service efficiency and provision of better services to the populace. It has been recognized that the failure of governments to embrace emerging ICT technologies could lead to severe disadvantages. Cloud computing has emerged as one of the ways of improving the ways that government institutions render their services [6]. Cloud computing has helped to address most of the challenges of implementing e-government and has modernized e-government systems through efficient use of resources and cost-saving [7, 8].

Many governments worldwide are already introducing or migrating to cloud computing as a means of improving their information technology (IT) governance, customization of services, reduction of costs, and enhancing the overall performance and quality of the public sector organizations [4, 7]. Several case studies have shown that most organizations are turning to cloud computing, including the public sector where it has already been adopted in several governmental institutions [4]; for instance, in the USA, the information portal of the federal government (USA.gov) has migrated its services to the cloud to solve the issue of huge network traffic, service delivery delays, long downtimes, and other issues facing them previously. The cloud migration reduced their site upgrade from 9 months to just one day; a 99.9% reduction in downtime and 72% reduction in annual cost were also reported [9].

Although numerous studies have been conducted on cloud computing adoption, a bibliometric study on the adoption of cloud computing in e-government is yet to be conducted. Therefore, this work aims to study the trends and developments that could aid the understanding of cloud computing adoption in e-government.

2 Methodology

This study employs a bibliometric analysis method using Biblioshiny for Bibliometrix in R-packages [10] to map the research literature on the adoption of cloud computing and e-government articles published between 2009 and February 2021. The articles were downloaded from the SCOPUS database as it is the largest database of peer-reviewed articles globally, covering mainly publications and journals [11]. Bibliometrics is the study of the quantitative aspects of scholarly publications, dissemination, and the use of citation information by the application of mathematical and statistical methods to papers, books, and other resources [12]. The strings of the research include the keywords “Cloud Computing “OR” Cloud Computing Adoption” AND “Government” OR “E-Government” OR “Public sector” for the retrieval of the published articles from SCOPUS. In this reference interval, 157 citable documents published between 2009 and 2021, including 47 articles, 77 conference papers, 27 book chapters, 2 books, 2 review papers, and 1 note were retrieved. The method of this study is similar to [13]; the flow diagram in Fig. 1 shows the employed search strategy in the current research.

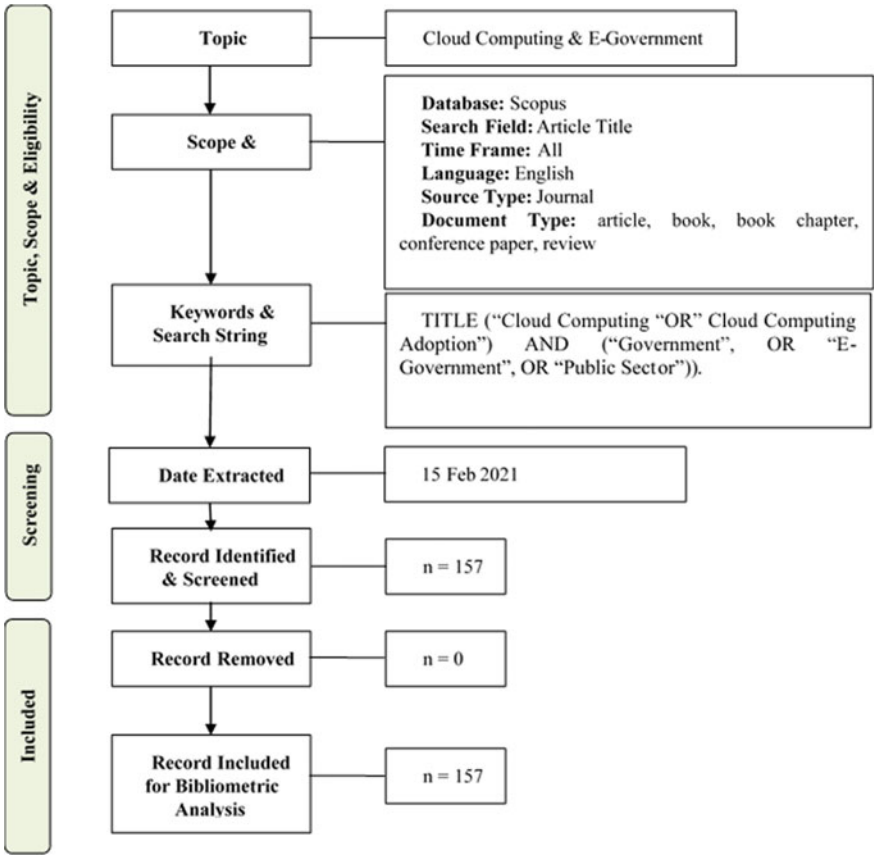


Fig. 1 Flow diagram of the search strategy

3 Results

The results of the analysis are presented in this section based on the indicators proposed in the methodology. Figure 2 shows a summary of the main information of the collected Bibliometric data.

3.1 Number of Publications by year

In Fig. 3, the indicators show a growing interest in the topic of cloud computing adoption and e-government since 2009, with 2015 being one of the most productive years (24 published articles), followed by the year 2011 (23 published articles).

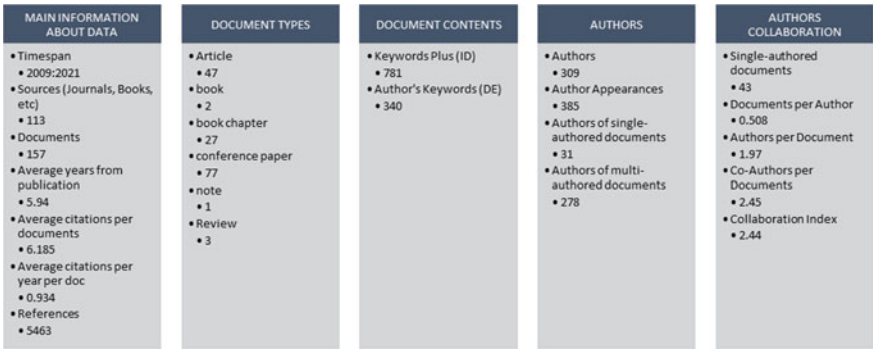


Fig. 2 Summary of the main information of collected bibliometric data

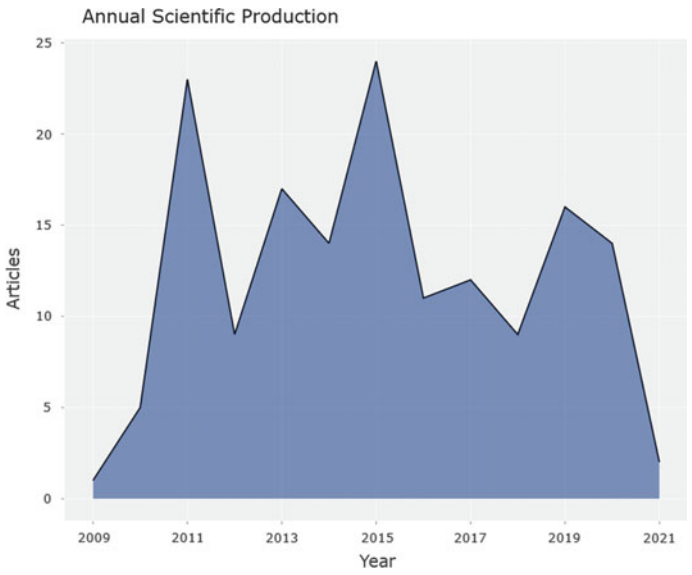


Fig. 3 Number of publications by year

3.2 Most Productive and Cited Countries

The most cited and productive countries in the areas of e-governance and cloud computing were identified using the "Country Scientific Production" as an analysis method in the Biblioshiny for Bibliometrix in R. The minimum and maximum numbers of documents per country were set at 2 and 1, respectively. The top 10 countries with the highest number of publications on cloud computing adoption and e-government are shown in Table 1. China leads in the publication aspect with 21 publications, followed by Malaysia (8), UK (6), Korea (5), Australia (3), Greece

Table 1 Top 10 influential countries in adoption of cloud computing and e-government publications

Country	Articles	Ranking	Total citations	Ranking
China	21	1	84	3
Malaysia	8	2	92	2
United Kingdom	6	3	67	4
Korea	5	4	34	6
Australia	3	5	46	5
Greece	3	6	126	1
Indonesia	3	7	5	9
Brunei	2	8	3	10
Czech Republic	2	9	22	7
Egypt	2	10	14	8

Table 2 Most used author keywords

Words	Frequency
Cloud computing	108
e-Government	42
Public sector	12
Adoption	9
g-Cloud	6
Government	6
Framework	5
Challenges	4
Innovation	4
Institutional theory	4

(3), Indonesia (2), Brunei, Czech Republic, and Egypt (2). Considering the number of citations per country, Table 2 also showed that Greece has the highest number of total citations (126), followed by Malaysia (92), China (84), the UK (67), Australia (46), Korea (34), Czech Republic (22), Egypt (14), Indonesia (5), and Brunei (3).

Based on the initial analysis of the extracted studies from the Scopus database, most of the previous studies in cloud computing adoption and e-government services were conducted in the developed countries, with no focus on developing countries.

3.3 Most Used Keywords

The bibliometric mapping analysis suggested that the most used author keywords in the analyzed studies were “cloud computing” (f = 108), “e-government” (f = 42), and “Public sector” (f = 12). Table 2 illustrates the most used author keywords in

Table 3 Top ten most influent authors in terms of publications and citations

Paper	Total citations	TC per year
Zissis and Lekkas [14]	124	11.2727
Cellary and Strykowski [15]	63	4.8462
Mohammed et al. [16]	49	8.1667
Clohessy et al. [17]	48	6
Ali et al. [18]	27	4.5
Mohammed et al. [7]	26	5.2
Al-Ruithe et al. [19]	26	5.2
Wu et al. [20]	24	4
Zhang and Chen [21]	24	2
Lnenicka and Komarkova [22]	22	7.3333

the analyzed studies. These results would assist researchers in finding a sufficient number of studies on the adoption of cloud computing and e-government by referring to the aforementioned keywords.

3.4 Most Cited Authors

The top 10 most influential authors in terms of publications and citations in the adoption of cloud computing and e-government are shown in Table 3. The article that received the highest citation, "Securing e-Government and e-Voting with an open cloud computing architecture" was published in the Government Information Quarterly journal, in 2011. It received a total of 124 citations and was the most impactful article based on the citation per year (11.27 citations/y).

4 Discussion, Conclusions, and Future Research Agenda

Through the wide time-span Scopus database, this study looked for all the published articles on the adoption of cloud computing in e-government. The literature review showed that only a few articles have focused on cloud computing adoption in e-government.

Studies on the adoption of cloud computing in the public sector institutions, including governments remain preliminary as most of the reported works in this aspect focus more on the technical perspective of cloud computing and the potential of saving cost through its adoption. Frameworks and models have been proposed by several researchers for cloud computing in various e-services; For instance, the majority of the existing studies have discussed the potential benefits and possible drawbacks of adopting cloud computing in e-government. The researcher in a

previous study published early between 2017 and 2020, (e.g. [4, 7, 23, 24]) reviewed the literature of the models which were proposed for e-government services migration to the cloud. The review analyzed and classified the models into two different types. Statistics showed that 88% of the models in the literature were proposed from the software engineering perspective (Step-based, layered-based, and component-based models), while only 12% of the models were from the information systems perspective (Conceptual/Theory-based models). Hence, it was concluded that studies are yet to empirically investigate cloud computing adoption in e-government. From the context of developing countries like Iraq, Abd et al. [25] surveyed the relevant prospects of the Iraqi population for cloud computing services and e-Government and found that the qualified Iraqi citizens have proposed 112 different e-services. Hence, this study could be important towards the improvement of e-government services based on cloud computing in Iraq. Li et al. [26] in their study argued that cloud computing can efficiently solve the problem of insufficient information resource sharing, reducing network risks, and improve the efficiency of e-government use with flexibility, scalability, reliability, and low cost. Wu et al. [20] focused on the decision processes of public cloud computing adoption in e-government; the study developed a theoretical model and the related hypotheses for the evaluation of the data gathered from 227 Chinese public sectors. From the results, the three decision-making modes (IaaS, SaaS, and PaaS) on adoption had different determinants. For IaaS adoption, the determinants were technical & business factors while the determinants of PaaS adoption were technical & management factors. The determinants of SaaS adoption were technical, management, and business factors.

Shin [27] examined cloud computing services adoption in public sectors by considering the key factors that affect the behavioral intent. The factors influencing user perception of cloud computing were explored by theorizing a model of acceptance. The technology acceptance model (TAM) was modified by applying the Theory of Reasoned Action (TRA) to propose a new model for cloud computing acceptance. Accordingly, the model combined TAM with some influencing factors like availability, reliability, accessibility, and security. These factors are driven by considering security, benefits, accessibility, and availability as the constructs that enhance the prediction of user acceptance of cloud computing techniques. For the empirical evaluation of the proposed model, the perception of users in public firms was evaluated using the proposed model and the results suggested the influence of perceived features of cloud services on the user behavior and intention. Mohammed et al. [23] investigated the factors influencing cloud computing adoption in the public sector in Yemen. The Fit-Viability Model and Diffusion of Innovation Theory models were adopted by the study to propose the adoption model of the study. Data was collected from 296 IT workers in the public sector in Yemen. According to Al Mudawi et al. [24], the implementation of e-government in the public sector still faces many problems, with the main problems being lack of finance, poor IT infrastructure, and poor data security. This study focused on the issues facing the implementation of e-government cloud service models from the perspective of developing countries. The outcome of this study showed that the

factors influencing e-government cloud service models in developing countries are mainly those related to how the top management perceives cloud computing adoption. The other factors include the level of readiness in the organization, such as the available IT infrastructure, social trust in the adoption of new technologies, and internet availability. It is therefore important to highlight the benefits of adopting cloud-based e-government services in developing countries. A study has previously reviewed the influence of cloud computing adoption on e-government services sustainability [4] and summarized that cloud computing can meet the aspirations of public organizations of having a strategic way of handling the dynamicity of today's business environment, as well as deal with the issues of technological innovation and environmental pressures in their public business environment.

References

1. AlAjmi Q, Al-Sharafi MA, Chellathurai GJ (2021) Fit-viability approach for E-learning based cloud computing adoption in higher education institutions: a conceptual model. In: Al-Emran M, Shaalan K (eds) Recent advances in technology acceptance models and theories. Springer International Publishing, Cham, pp 331–348
2. Al-Sharafi MA, Arshah RA, Abu-Shanab EA (2017) Factors influencing the continuous use of cloud computing services in organization level. In: Proceedings of the international conference on advances in image processing 2017, Association for Computing Machinery, Bangkok, Thailand, pp 189–194
3. Al-Sharafi MA, Arshah RA, Abu-Shanab EA (2019) Questionnaire development process to measure the SMEs' continuous use behavior towards cloud computing services. In: Proceedings of the 2019 8th international conference on software and computer applications—ICSCA '19, 2019, Association for Computing Machinery, Penang, Malaysia, pp 50–55
4. Mohammed F et al (2020) The effect of cloud computing adoption on the sustainability of e-government services: a review. Int J Adv Sci Technol 29(5):2636–2642
5. Al-Sharafi MA (2014) A review of the factors affecting the success/failure of E-government projects. Saba J Inf Technol Netw (SJITN) 2(1)
6. Alajmi QA et al (2018) The effectiveness of cloud-based e-learning towards quality of academic services: an Omanis' expert view. E-learning 9(4)
7. Mohammed F et al (2017) Cloud computing fitness for e-Government implementation: Importance-performance analysis. IEEE Access 6:1236–1248
8. Qasem YAM et al (2021) Towards the development of a comprehensive theoretical model for examining the cloud computing adoption at the organizational level. In: Al-Emran M, Shaalan K, Hassanien AE (eds) Recent advances in intelligent systems and smart applications. Springer International Publishing, Cham, pp 63–74
9. Kundra VJFCI (2010) State of public sector cloud computing
10. Aria M, Cuccurullo C (2017) Bibliometrix: an R-tool for comprehensive science mapping analysis. J Inform 11(4):959–975
11. Aghaei Chadegani A et al (2013) A comparison between two main academic literature collections: Web of Science and Scopus databases. Asian Soc Sci 9(5):18–26
12. Leung XY, Sun J, Bai B (2017) Bibliometrics of social media research: a co-citation and co-word analysis. Int J Hosp Manag 66:35–45

13. Zakaria R et al (2021) Worldwide melatonin research: a bibliometric analysis of the published literature between 2015 and 2019. Chronobiol Int 38(1):27–37
14. Zissis D, Lekkas D (2011) Securing e-Government and e-Voting with an open cloud computing architecture. Gov Inf Q 28(2):239–251
15. Cellary W, Strykowski S (2009) e-government based on cloud computing and service-oriented architecture. In: Proceedings of the 3rd international conference on Theory and practice of electronic governance, Association for Computing Machinery, Bogota, Colombia, pp 5–10
16. Mohammed F, Ibrahim O, Ithnin N (2016) Factors influencing cloud computing adoption for e-government implementation in developing countries: instrument development. J Syst Inf Technol 18(3):297–327
17. Clohessy T, Acton T, Morgan L (2014) Smart city as a service (SCaaS): a future roadmap for e-government smart city cloud computing initiatives. In: 2014 IEEE/ACM 7th international conference on utility and cloud computing
18. Ali O, Soar J, Yong J (2016) An investigation of the challenges and issues influencing the adoption of cloud computing in Australian regional municipal governments. J Inf Secur Appl 27–28:19–34
19. Al-Ruithe M, Benkhelifa E, Hameed K (2017) Current state of cloud computing adoption—an empirical study in major public sector organizations of Saudi Arabia (KSA). Proc Comput Sci 110:378–385
20. Wu J et al (2016) Investigating the determinants of decision-making on adoption of public cloud computing in e-government. J Glob Inf Manage (JGIM) 24(3):71–89
21. Zhang W, Chen Q (2010) From E-government to C-government via cloud computing. In: 2010 International conference on e-business and e-government
22. Lnenicka M, Komarkova J (2019) Developing a government enterprise architecture framework to support the requirements of big and open linked data with the use of cloud computing. Int J Inf Manage 46:124–141
23. Mohammed F et al (2017) Cloud computing adoption model for e-government implementation. Inf Dev 33(3):303–323
24. Al Mudawi N et al (2020) Issues and challenges: cloud computing e-government in developing countries. Int J Adv Comput Sci Appl 11(4):7–11
25. Abd T et al (2019) Iraqi e-government and cloud computing development based on unified citizen identification. Period Eng Nat Sci 7(4):1776–1793
26. Li Y, Zhu L, Tu W (2019) Research on E-government data management in cloud computing environment. In: 2019 International conference on smart grid and electrical automation (ICSGEA)
27. Shin DH (2013) User centric cloud service model in public sectors: policy implications of cloud services. Gov Inf Q 30(2):194–203

Implementing Knowledge Management in an IT Startup: A Case Study

Dena Ahmed, Said A. Salloum, and Khaled Shaalan

Abstract Knowledge management (KM) comprises of managing, planning, deploying, collecting, storing, reusing, and distributing the knowledge in organizations in an organized planned way. Applying KM in an organization aims to allow for ensuring overall success as it is a method that simplifies the entire process of dealing with knowledge. Some organizations such as software development companies suffer from many issues that obstruct the KM process. This case study performed on an Egyptian IT startup aims to examine their business process and methodologies for dealing with knowledge. The results showed that they were wasn't utilizing the knowledge they had properly and suffered from a lack in KM in almost all aspects. All problems were examined and solutions to handle them were proposed to implement efficient KM.

Keywords Knowledge management (KM) · Knowledge management system (KMS) · Software engineering (SE) · Software development · Startup

1 Introduction

Software development is quite elaborate involving several people working in multiple stages and different activities. New technologies are constantly emerging, and different projects require different development techniques. Therefore, many new issues are faced and new experience is gained with each project and thus, new knowledge is created. That is why knowledge in software development is very

D. Ahmed · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_59

heterogeneous and diversified with an exponential growth rate [1, 2]. Companies, especially startups, face many difficulties while trying to keep track of what the knowledge they have is, where it is located, who has it, and how they can apply it. This further proves the need for some structured methodology to manage the knowledge at hand, helping the organization to advance forward. KM has a lot to offer to software companies as it takes an all-encompassing, exhaustive approach to the different activities of collecting and handling knowledge [3]. It considers the knowledge obtained from the software developers' minds as human assets and converts that into a company asset that can be used and shared by all the company's employees [2]. Having a KM system in a software company creates a shared language between the software developers, which facilitates their interactions with each other, thus allowing for more sharing of their experiences and knowledge. KM simplifies and advances the concept of organizational learning in the context of software development by helping with knowledge transfer throughout the organization [4], thus allowing for more success and improved performance. Many companies, especially startups, are unaware of how KM can immensely impact their quality of work and productivity [5], thus losing the chance of advancing. Several startups go out of business because of the lack of proper KM practices [6]. This case study aims to take a closer look at an Egyptian IT startup to identify the aspects that could benefit from applying KM, and show how important KM is to the success of software organizations.

The following points will be tackled by this study:

(1) What are the various types of knowledge processes used in the company?
(2) How important is applying KM in a software startup?
(3) What are the problems regarding KM in the company?
(4) What are some solutions and plans for applying proper KM?

This paper is divided as follows. Section 2 discusses KM in software organizations. Section 3 covers the methodology of data collection for the case study. Key issues, KM processes identified, and proposed potential solutions are presented in Sect. 4. Finally, Sect. 5 will conclude.

2 Knowledge Management in Software Organizations

Knowledge in SE is very dynamic and continuously evolving with new technologies developed every day, and so does the organization's culture. The processes in SE [7] are almost wholly dependent on the knowledge of the developers and team members. The fundamental principle in SE [8] is that the availability and efficient utilization of knowledge significantly enhances the quality of the developed software products. Multiple factors assist the KM process in software companies such as technology, leadership and organizational culture [9]. Technology connects developers together and creates a knowledge bank for the entire company; while

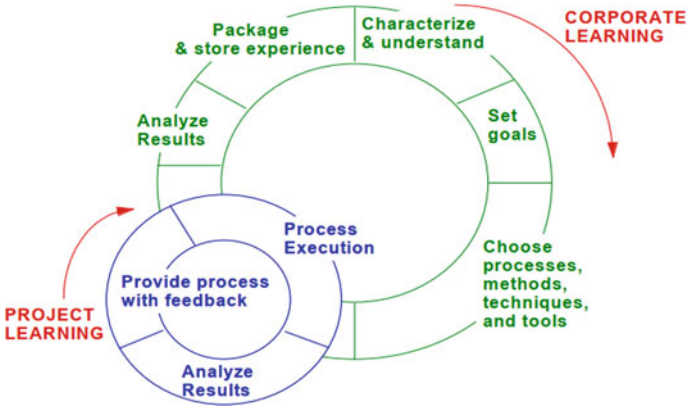


Fig. 1 The quality improvement paradigm

leadership is the force that pushes for applying KM in the work processes and services of the company. The organizational culture advocates knowledge sharing and distribution of expertise to advance innovation within the company [10]. There is a continuous learning process occurring during the lifecycle of software development, therefore all knowledge from projects must be appropriately gathered and documented in an organized manner to help with upcoming projects. This concept is also portrayed in the Quality Improvement Paradigm (QIP) [11, 9] where analysis and finding conclusions regarding the project's progress is done at each project level. Then, the results are further analyzed up another level and stored in what can be referred to as an experience database. This serves as knowledge to guide plans for upcoming projects such as selecting the business processes, methodologies and tools to be used [12]. Figure 1 shows the QIP flow.

3 Methodology

The startup under study will be referred to with a pseudo name "Tech-K" throughout the paper for confidentially purposes. Data about Tech-K was collected through interviews with all employees. Questions for the interviews were designed carefully to address the research variables to identify the degree of KM implementation in an IT startup.

Tech-K is an IT startup that started with 5 employees, and now currently 20 employees: 2 Android developers, 2 iOS developers, 2 front-end Web developers, 1 backend developer, 4 employees handling social media and marketing tasks, 2 account managers, 2 UI designers, 3 interns, 1 technical officer and the CEO. In the beginning, the focus was only on developing websites. This later expanded into developing basic mobile applications using drag & drop app builders, then native

app development using Android then later iOS. In addition, social media marketing was also added to the company's activities. The CEO has a medical and business background, so his technical knowledge regarding software development is limited. Almost all employees are young, fresh graduates with limited experience.

4 Results and Discussion

4.1 Identified KM Processes and Issues

Acquiring Knowledge. Because of the small number of employees, meetings are usually informal where many different subjects are discussed. The outcome of these meetings is not being utilized at the organizational level so there is no proper attainment of knowledge to keep track and predict further developments. Experiences shared during meetings where developers discuss their progress, issues faced and plan for remaining tasks are the main source for acquiring knowledge but they are not recorded to benefit from. Requirements' gathering from clients is another source of gathering knowledge but it is also done informally through meetings and interviews without proper documentation or an organized procedure. When more developers join the company, the CEO and technical officer are going to be the only source of information. This knowledge centralization forces the technical officer to have multiple sessions and meetings to transfer their knowledge instead of actually focusing on more important pressing tasks. So, the CEO and the technical officer are subjected to loss of productivity because of the centralization of knowledge.

Codifying Knowledge. The majority of the employees don't know where to find knowledge about activities or projects done before or even the current knowledge regarding ongoing projects. Knowledge is not well recorded in most cases. A developer wanting to know about the logic of a specific feature in a mobile application has no documentation to go back to; but has to ask the technical officer or the UI designer. They don't have the answers to all the questions and they have to go back to ask the client sometimes. Occasionally, assumptions are just made without knowing if that assumption will match the client's needs. Employees think that there has to be a specific person responsible for documenting project requirements in a proper way, where they are continuously updated with any changes or modifications for an enhanced development environment.

Disseminating Knowledge. Because most of the knowledge is distributed through informal conversations, it is harder to supervise or control. Employees are of the opinion that informal conversations are used because there is not a better alternative. When knowledge is being recorded in an accessible system in a user-friendly way, clearly demonstrating who is working on what and which individual has the knowledge regarding a specific point, then the dependency on the informal meetings becomes less essential.

Almost all employees agree that a big problem at Tech-K is that the knowledge is stored only in the minds of the employees without being properly recorded. There is no doubt that this implicit knowledge is a great asset, but it would be of much greater benefit if the company was somehow less dependent on the actual presence of those employees. This is because this knowledge is lost when an employee leaves Tech-K which causes significant problems to the workflow, especially with being a startup with limited resources where each employee is handling important tasks that can't be delayed. There isn't a proper KMS to solve this. Hiring experienced senior developers in a startup is challenging as they need competitive salaries while there are limited financial resources. Senior developers also have more experience so they have worked with different technologies and bigger companies with better organization. So, even if offered a high salary, they usually leave after a few months because of the lack of a proper organizational structure. When a senior developer is working on a project, and then they leave the company before finishing it, all of his knowledge is not recorded. Even with a handover process, the complete knowledge inside the senior developer's head would be lost. Also, with no proper documentation of the code architecture, approach or techniques used in the code, the developer handling the rest of the project will face a lot of difficulties. This cripples the startup as the remaining junior developer would struggle with continuing the project on his own, and hiring another senior developer would definitely take a lot of time that the company can't afford. Some employees think that sharing knowledge with others will strengthen their standing in the company since you are likely to get back knowledge from others if you share your knowledge with them. Others at Tech-K finds sharing knowledge with colleagues to be very threatening. This is the case usually between developers or employees of the same level.

Developing & Applying Knowledge. It was found that at several times, the same knowledge is being developed at more than one place at Tech-K. This happens because of the absence of detailed information about colleagues' activities. For example, both front-end web developers would be working on a specific task at the same time because of some miscommunication or the lack of proper documented task allocation. This costs time and money, which is crucial for a startup. Developers also complain that they have to wait and keep asking for the designs from the UI developers and that slows down their progress. With the limited number of only 2 UI designers, they are under a lot of pressure and face difficulties in delivering their tasks on time. With the lack of resources and a prioritized task allocation system, there are a lot of informal requests. The same issue happens with the back-end developer as he is handling all web and mobile projects. Another complaint by the mobile developers was that the UI designers were too ambitious with their designs. The UI team always tries to create overly advanced designs with complicated patterns in an attempt to showcase their skills and impress both the client and the CEO. However, these designs are most of the time complicated for junior developers to translate into code and they take up a lot of time that could have been better utilized on coding main feature. Another issue is the constant change in requirements. Since most projects are obtained through

personal connections of the CEO, there are not strict guidelines regarding the requirements of the project. For example, if the client wants a mobile application, the contract would only include the general broad functionality of the app without details. Also, in the spirits of being more flexible and satisfying clients to the best possible way, the CEO doesn't enforce any regulations regarding requirements change to maximize customer satisfaction. However, that caused many significant issues through the development process. Giving the client the freedom to constantly change the requirements whenever they want opened the door to an endless loop of modification requests that didn't make sense most of the time. Most of the time, clients don't have enough technical knowledge so they are not good judges of what is applicable and its impact on the overall project. There are several occasions where the client is confused and doesn't know what they want, so they come up with non-feasible unnecessary requirements. This affects the development process negatively in many ways such as delaying the deadlines affecting other future projects, having to destroy or completely change code architecture based on earlier requirements, changes across several departments and many more. Moreover, sometimes the CEO approves modifications without understanding the technical effort that would be needed by the developers to accommodate those changes. Some changes can completely change the entire structure and architecture serving as the building layer of the entire project which would require extensive effort and time.

In addition, there are no dedicated testers to test the projects before submitting them to the customer. The developers & other employees play the tester role by trying different scenarios on the product to identify bugs. Some mobile apps require testing on devices with different specifications, operating systems and screen sizes; and that is hard because of the limited resources available at Teck-K. Thus, employees test the apps on their personal mobile phones. Having no dedicated testers result in poor quality products and unsatisfied clients because of bugs or crashes.

Knowledge Systems. Teck-K uses Slack (*Slack* n.d.) for communication between employees. Employees don't feel like it is sufficient for communication and prefer a more comprehensive platform with more focus on project management. Developers also use Git (*Git* n.d.) as a version control system to monitor modifications in a project among several developers. It also allows developers to push their code changes to a centralized remote repository on the server. Keeping track of the history of changes made to all files in the code allows going back to a previous working version in case of bugs, and helps in identifying who is the responsible party if needed. However, developers don't utilize Git correctly. Most of them don't regularly push their new changes to the remote repository, and when they do, they don't describe the changes done. This also causes conflicts between developers in case a developer made some changes on his machine locally but didn't push to the remote copy, and another developer continued working on the old version without those updates.

Issues Summary. Issues faced by Tech-K can me summarized to: Not learning lessons from previous projects, flawed team selection and task allocation, centralization of knowledge, lack of important knowledge among key employees, lack of

proper knowledge sharing and transfer, reluctancy of some employees to share knowledge, staff turnover causing loss of knowledge/failure of projects, poor requirements gathering from customers, lack of a proper change control system for modifications in requirements, project management problems including poor management skills, lack of proper KMS, misuse & underutilization of systems used, scheduling and deadline estimation problems, lack of proper testing of software products, and lack of a proper knowledge map. All of these issues cause significant damage to the startup including unsatisfied clients, frustrated exhausted employees, poor quality products, delayed delivery, poor reputation and loss of money. If these problems are not properly addressed, Tech-K won't be able to grow and would eventually fail.

4.2 Proposed Potential Solutions

Create a Knowledge Sharing Culture and a proper KMS. A knowledge repository similar to the portal in [13] is recommended. This system should contain all the information, expertise, help, guides and support needed for all tasks' domains. This could include IDE custom settings, authentication keys, libraries used, frequently encountered issues with documented solutions. Template code files can also be included for commonly used tasks across several apps and complex code components can be documented to be reused instead of creating everything from scratch. This saves effort, money and time by reusing the previously gained knowledge for more efficient advancement. This will promote the knowledge sharing culture at Tech-K and encourage employees to share their knowledge since they are benefiting from it as well.

Implement a Change Control/Management System. With the rapid and continuous change in technology and increase in competition, clients are not sure of what they need in a software product [14], resulting in many unreasonable and often unnecessary requirements. Therefore, this process has to be controlled by a proposed requirement change model. The process should start with creating a "change request", then Tech-K would start "assessing the change", where developers analyze the change request, identify the system components to be changed, estimate the cost required for this change and pinpoint potential conflicts with other requirements. The next step is "deciding on the change", where a decision is reached whether the change will be implemented or not depending on the knowledge assessed including rationale, relationships between requirements and design components, and impacts of change on the development process. If the changes are approved, they move to the next stage "planning the change", where the developers divide the new modifications into tasks and allocate them to responsible parties. At last, in "executing the change" stage, the changes are executed and then further validated after completion.

Implement Proper Testing/Quality Control. As discussed earlier, Tech-K doesn't have dedicated testers and the employees test the software products

themselves before submitting to the clients. This results in undiscovered bugs or crashes leading to problems with the client. The proposed solution is to hire dedicated testers and apply a comprehensive testing methodology where all identified bugs are categorized with detailed steps to reproduce and actual vs expected behaviors thoroughly documented. Bug fixes should be updated and documented as well with multiple testing cycles. This will improve the quality of the products and enhance the reputation of Tech-K.

Usage of Knowledge Management Systems/Tools. UI designers at Tech-K used Adobe Photoshop to create full designs, then cut individual graphical assets with the different sizes and specs, then send them to the developers through email or using a flash drive which is not optimal by any means. Using Zeplin (*Zeplin* n.d.), they can turns designs into specifications and related snippets of code corresponding to design components for each screen, with the individual graphic components such as icons, fonts, color schemes, different sizes and many more are ready for the developers to download and use. Slack was used at Tech-K, but it was not efficient and didn't need the required needs. The proposed tool is Basecamp (*Basecamp* n.d.), which is a project management application that provides all tools to create to-do lists, message team members, create schedules, upload documents, check-in and create groups/projects. It enhances collaboration and work flow management by organizing communications and projects. Different projects can be created where all related information is saved including full documentation, any changes in requirements, deadlines, milestones, task allocations to developers, client feedback, to-do lists, testing results, and many more. Different teams can also be created where team members can communicate easily and share related knowledge. Schedules, announcements and many other features are also present. This helps in keeping all the information and knowledge about projects in an organized way that will always be available as assets that employees can always go back to.

Currently, the back-end developer at Tech-K creates the API requests and sends the links of each request through messages on Slack. This is a poor technique since they can't be tested and are hard to find later. The proposed tool is Postman (*Postman* n.d.), which is a collaborative tool for creating RESTful APIs for the projects, allowing the back-end developer and other developers to test the requests/responses of the links to make sure they are in the required format and working as expected. It is much easier than having to write separate unit tests to confirm that the API is working correctly. It enables the organization of APIs and it has a user-friendly user interface making it easy to use. Moreover, developers need proper training for Git usage. Detailed commit messages describing the changes done in the code should be added. Also, bulk commits with unrelated files should be avoided, and only related files with relevant modifications should be included in the same commit. This will enable better tracking of changes and makes looking at the files' history easier.

5 Conclusion

This study demonstrated the usage, benefits and impact of KM in SE based on a case study conducted in an Egyptian IT startup, Tech-K. The activities in the lifecycle of software development were examined. The effect of knowledge sharing, culture, technology, expertise, leadership, resources' utilization was also evaluated. The results show that the techniques and methodologies currently used at Tech-K were poor and inadequate to apply KM in an effective manner that would benefit the company. Several problems were identified including not learning from previous mistakes in older projects, lack of knowledge sharing, staff turnover leading to loss of knowledge, poor requirements gathering, lack of a change control system, poor documentation, lack of testing leading to poor quality software products and generally a lack of a proper knowledge map. All of these issues affect the performance of Tech-K negatively, threaten their presence in the market and stand in the way of their growth or stabilization. Potential solutions were proposed to apply KM properly at Tech-K. The first solution is to create a knowledge repository to encourage knowledge sharing and reuse the existing knowledge by documenting frequently encountered errors with solutions, templates, external libraries, code snippets and other useful knowledge sources. Another solution is implementing a change control system to manage the non-stop change of requirements from the clients and reign it in to avoid any negative impact on the projects or the company, thus making sure that only the necessary and beneficial requirements are approved. Implementing a proper testing methodology is also proposed by hiring a dedicated tester and performing clear testing documentation with all faced bugs and issues. The last proposed solution was using some tools and applications to facilitate the flow of knowledge at Tech-K such as Zeplin, Basecamp and Postman. Applying these solutions will significantly improve KM at Tech-K, enhance the quality of software products and boost the entire company's performance leading it to the path of becoming a successful stable organization.

References

1. Foote A, Halawi LA (2018) Knowledge management models within information technology projects. J Comput Inf Syst 58(1):89–97. https://doi.org/10.1080/08874417.2016.1198941
2. Rong G, Liu X, Gu S, Shao D (2018) A goal-driven framework in support of knowledge management. In: Proceedings—Asia-Pacific software engineering conference, APSEC, 2018, vol 2017 Dec, pp 289–297. https://doi.org/10.1109/APSEC.2017.35
3. Razzak MA, Mite D (2015) Knowledge management in globally distributed agile projects—lesson learned. In: Proceedings—2015 IEEE 10th international conference on global software engineering, ICGSE 2015, pp 81–89. https://doi.org/10.1109/ICGSE.2015.22
4. Ochoa SF, Robbes R, Marques M, Silvestre L, Quispe A (2017) What differentiates chilean niche software companies: business knowledge and reputation. IEEE Softw 34(3):96–103. https://doi.org/10.1109/MS.2017.64

5. Pham QT, Nguyen DT (2017) An empirical investigation of knowledge management in Vietnamese SMEs. In: Proceedings of the 2017 17th international conference on computational science and its applications, ICCSA 2017. https://doi.org/10.1109/ICCSA.2017.8000016
6. Zafar I, Shaheen A, Nazir AK, Maqbool B, Butt WH, Zeb J (2018) Why Pakistani software companies don't use best practices for requirement engineering processes. In: 2018 IEEE 9th Annual information technology, electronics and mobile communication conference, IEMCON 2018, 2019, pp 996–999. https://doi.org/10.1109/IEMCON.2018.8614913
7. Riaz MN, Buriro A, Mahboob A (2019) The effect of software development project team structure on the process of knowledge sharing: an empirical study. In: 2019 2nd International conference on computing, mathematics and engineering technologies, iCoMET 2019. https://doi.org/10.1109/ICOMET.2019.8673504
8. Al Hafidz MU, Sensuse DI (2019) The effect of knowledge management system on software development process with scrum. In: ICICOS 2019—Proceedings of 3rd international conference on informatics and computational sciences: accelerating informatics and computational research for smarter society in the era of industry 4.0. https://doi.org/10.1109/ICICoS48119.2019.8982506
9. Vasanthapriyan S, Tian J, Xiang J (2015) A survey on knowledge management in software engineering,. In: Proceedings—2015 IEEE international conference on software quality, reliability and security-companion, QRS-C 2015, pp 237–244. https://doi.org/10.1109/QRS-C.2015.48
10. Rodrigues LLR, Mathew AO (2019) Knowledge management technology, knowledge sharing and learning—a case study. In: 2019 International conference on automation, computational and technology management, ICACTM 2019, pp 273–276. https://doi.org/10.1109/ICACTM.2019.8776779
11. Jahan MS, Talha Riaz M, Arif KS, Abbas M (2019) Software project management and its tools in practice in IT Industry of Pakistan. In: 2019 2nd International conference on computing, mathematics and engineering technologies, iCoMET 2019. https://doi.org/10.1109/ICOMET.2019.8673535
12. Ouriques R, Britto R, Wnuk K, Ouriques JF, Gorschek T (2019) A method to evaluate knowledge resources in agile software development. In: International symposium on empirical software engineering and measurement, vol 2019, Sept 2019. https://doi.org/10.1109/ESEM.2019.8870167
13. Vasanthapriyan S, Tian J, Zhao D, Xiong S, Xiang J (2017) An ontology-based knowledge sharing portal for software testing. In: Proceedings—2017 IEEE International conference on software quality, reliability and security companion, QRS-C 2017, pp 472–479. https://doi.org/10.1109/QRS-C.2017.82
14. Chaisricharoen R, Srimaharaj W, Boonyanant P, Laohapensaeng T (2019) Virtual workspace for a small-scale software development company in developing countries case of Thailand. In: International symposium on wireless personal multimedia communications, WPMC, vol 2019, Nov 2019. https://doi.org/10.1109/WPMC48795.2019.9096187

Managing Knowledge in Social Media

Mohamed Abdulla, Omar Khalid, Said A. Salloum, and Khaled Shaalan

Abstract Social media platforms are massive sources of information and knowledge, considering their nature that heavily relies on the concept of media sharing in the form of visuals, written, or other multimedia formats. The ease of accessing these channels and contributing to them is what makes it a very unique yet challenging source to manage. The variety of individuals and their backgrounds, along with the numerous numbers of active users that is rapidly increasing, makes it very difficult to practice basic and traditional KM processes. This systematic review highlights some of the initiatives that have promisingly attempted to acquire the knowledge in SM, verify it, and create a reliable source out of it. While many case studies have proven that the knowledge exists in today's social platforms can be of a great benefit, very least has been done to easily capture original, relative, and beneficial data from it. We reached to the finding that KM researchers focused on written materials, while neglecting platforms that provided audio/video based content and viral media platforms as well, which could be the richest platforms in containing genuine knowledge. Many notable efforts were attempted to apply frameworks that would effectively capture the knowledge in SM, verify it, and turn it into a reliable knowledge reservoir, and some of the experimental case studies had promising results, but the downfall was the use of traditional KM processes rather than utilizing technological or intelligent tools. We believe that there is a significant amount of data that is being wasted, which needs a technological software that could extract this data and turn it into a knowledge repository.

Keywords Knowledge management (KM) · Social media (SM) · Knowledge management tools · Knowledge management systems

M. Abdulla · O. Khalid · S. A. Salloum (✉) · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, United Arab Emirates
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Research Institute of Sciences and Engineering, University of Sharjah, Sharjah, United Arab Emirates

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_60

1 Introduction

Individuals nowadays spend numerous amounts of time on social media platforms tweeting, posting, reading, watching, and sharing a variety of contents. These different contents are considered to be a massive source of knowledge, information, and data. Social media and knowledge management have both been the center of many academic and organizational researchers' attention. However, managing the knowledge in social media has recently gained their interest giving the amount of knowledge present in these social spaces. This article highlights some of the distinguished attempts toward this field. Our goal is to give an overview of some of the actual and experimental studies that clearly underlined the challenges as well as the potentials of this particular area.

The study targets researchers in the field of knowledge management in social media and social platforms, general knowledge management researchers who are interested in new sources of knowledge and information and how to manage them, as well as business organizations that are seeking to improve their customer experience by utilizing these platforms [1].

This paper is structured as follows: the following subsections will give a basic introduction of knowledge management, social media, and knowledge management in social media. Section 2 will describe the research method, including papers selection mechanism, quality assessment, and research questions. Section 3 presents a systematic review of some of the well-recognized works in this field, pointing the strengths and weaknesses in each of them. Section 4 will show the outcomes of this study in the form of answers to the research question, and Sect. 5 will conclude.

1.1 Knowledge Management

Knowledge management is considered to be a critical strategic tool for organizations to achieve better performances and higher levels of competitiveness in an environment of frequent job-shifting behavior by employees [2]. Although there is no unanimous agreement yet on what constitutes knowledge management, in this paper, we will choose the definition of Davenport as he has defined knowledge management as "a method that simplifies the process of sharing, distributing, creating, capturing and understanding of a company's knowledge" [3]. In other words, we can describe knowledge management as Organizational Learning. According to Stata [4], organizational learning is different from individual learning in terms of it requiring shared insight, knowledge, and models. Besides, it does not rely only on the memory of the participants of the organization; rather, it depends on the whole culture of the company, i.e. (strategies, policies, defined models and explicit processes). The main two strategies of knowledge management are Codification and Personalization. Codification is to organize and store information that constitutes

the knowledge of the organization and make it available to all related parties in the organization. Personalization is having a centralized store of data to support the flow of the information within the organization.

1.2 Social Media

Social Media is "a group of internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of User Generated Content" [5]. It refers to a group of online services that provide social interactions for users as it allows them to evaluate, share, co-create, and find the online information repository. Another terminology that describes Social Media is Social Software, which is the term used to label the use of computing tools to generate added value from social activities. Social Software includes weblogs, photo and music sharing, instant messaging, message boards, and mailing lists [2]. There is an important distinction between traditional communication software, which aligns people in groups with a top-down approach like the case of forums and discussion lists, and the bottom-up approach Social Software which allows members to form networks based on their preferences [6].

1.3 Knowledge Management and Social Media

Knowledge Management in social media depends on a theory called social exchange theory, which is a behavior in social media that allows exchanging views and opinions. This theory is critical when it comes to KM because the content of the knowledge that's being exchanged is beneficial for the business itself [7]. That is why most IT organizations use social media platforms for the development process of their KM system, which makes it more useful. For example, Facebook allows an open graph feature that enables users to like and comment, which is an excellent example of social exchange theory. The method has a positive impact on the organization, which allows them to develop the KM process much more comfortable and making their content available for the employees.

Social media and knowledge management are on the emerging trend in our generation; it is estimated that social media could have a $1.3 trillion impact on business, according to McKibsey, most of which comes from productivity improvement among knowledge workers [7]. There are two fundamental approaches for KM in social media, which are storage based and communication-based, and the approach depends on what the technology the organization adapts, not how the way it works. For example, an Engineering firm that stores all the technical knowledge in the company's server, in which they are retrieved by the employees' needs, and employees can store future knowledge in those servers. On the other hand, a law firm that does not adapt the file-sharing technology, instead it shares

knowledge through direct communication (formal). However, if people did contribute knowledge to these systems, the amount of the content available will make it challenging to navigate through the reliable content.

There are a couple of tools that overcomes the challenges experienced in the previous generation. For example, the social media affordance of transparency makes the contribution available to one another, which basically captures the knowledge as it is embedded in practice rather than a preserved abstract. This transparency also allows employees to learn about who knows what in the organization without necessarily even realizing they are acquiring this knowledge; this phenomenon is called ambient awareness. Digital trace is widely applied in many social media platforms, which basically constructs a knowledge profile for employees based on the content of their interaction and makes that content searchable for others. Social media platforms algorithmically decide which content is relevant to the searcher through newsfeeds and other search mechanisms. The significant advantage social media offer is allowing access for people to KM anytime and anywhere resulting in "always-on" knowledge sharing.

Cloud-based social media changed a lot in the KM system, where it enables users to interact and access the organization cloud remotely and enables the user to upload/download content. Wikipedia allows editing for any given content just by clicking, and based on the Wikipedia environment; knowledge is being verified. Facebook, on the other hand, allows users to share status and content publicly and privately.

2 Research Method

In this study, we will attempt to perform and present a systematic review [8] of experimental studies and researches in the field of knowledge management in social media. The purpose of this study is to demonstrate an overview of the most distinguished efforts that have been made in this field, what criteria have been addressed, and how they have been addressed. In particular, this research will attempt to answer the following questions:

1. Which platforms are mainly being considered as a source of information/knowledge
2. How up-to-date are those platforms?
3. Are any of the current trending platforms being utilized, such as Twitter, Facebook, Instagram, YouTube, Snapchat, etc.?
4. What are the most platforms among them all that are actually holding some usable knowledge in its overall content?
5. What are the primary resources of knowledge in these platforms, and can these information be accessed or found much easier elsewhere?
6. How can we verify the originality of these information knowing that these platforms contain a lot of misleading or inaccurate information?

7. What has been done so far (any effort) to take advantage of these platforms in terms of knowledge management?
8. How can the spread or scattered knowledge in those platforms be combined or gathered in a usable, reliable, and accessible format?
9. Are there any suggested handy methods that have been widely used to gather information or knowledge?
10. Are there any tools or software that have been developed to efficiently extract decent knowledge from those platforms?
11. These platforms significantly vary in terms of content representation (i.e., Twitter is mainly written. Instagram is using photos and short video, whereas YouTube relies on long videos). How can this variety be handled?
12. Are there any actual case studies where an organization implemented some sort of social media KM and it has been proven to actually work?

The idea behind this study and its structure was initiated by identifying the trending research topics in the field of knowledge management in general. After compiling those trending topics, we started looking for an area that involves modern technology and can remain as an active field of study for a considerable amount of time. Moreover, we were looking for an area in the KM field where people have recently been paying very close attention to, but at the same time, very minimal overviews and studies have been provided. Despite the many previously published reviews in this file, there is still a considerable number of unanswered questions and unaddressed areas. Based on these vague areas, the above questions were developed. The motivation of this study is to present a fair assessment, a thorough overview, as well as a subjective evaluation of the quality of the studies in this particular field. After identifying the research questions that this study will be based on, a selection process has been developed to recognize the best available studies in terms of specific criteria that can effectively answer our questions. The coming sub-sections will address each of them in more detail.

2.1 Trending Topics in Knowledge Management

We started looking into the most recent studies and research papers that have been published in the area of knowledge management that involves some modern technologies. Among these areas, we came across many exciting topics such as: combining mobile technologies into knowledge management, knowledge visualization, KM user interface (UI) improvements, and KM in social media, including users' engagement, chatting tools, and digitalization. Subsequently, the collected research papers have been indexed as part of an inventorying procedure. The indexing practice involved a vast collection of highly reputed databases and resources, such as Google Scholars, Research Gate, Engineering Village and Science Direct. The various databases were thoroughly scanned based on predefined vital phrases that appeared in headings or as keywords. The search phrases

merged knowledge management, social media, technologies, frameworks, knowledge management tools, knowledge management systems… etc. The last topic has significantly caught our attention for many reasons: the vast number of active users, the incredibly rapid pace of development, and the significant amount of knowledge available on these platforms.

2.2 Developing Research Questions

There has been a considerable number of studies when it comes to the field of knowledge management in social media. Researchers have adequately examined most of its aspects, yet leaving many unanswered questions and vague areas behind. At the earliest stage of this research, we attempted to collect as much data as possible in this field and critically analyzing them. Then, we categorized these data into four main categories: social platforms that have been widely discussed in the field of knowledge management, the type and quality of information presented by these platforms, frameworks, and methods introduced to capture and manage the knowledge in those media, and triumphant stories and case studies. Based on the data collected, the above categories, and our initial curiosity, we came up with the above-mentioned research questions.

2.3 Selection Process

After identifying the articles, the first step was to discard duplicate papers and any paper that does not attempt to answer any of our research questions. We also excluded all outdated reviews that discuss knowledge management in platforms that have been discontinued knowing that the area of social media has and will always be a fast-growing and significantly changing area. Afterward, a quality assessment has been performed based on the number of citations, the originality of the content, the ranking of the journal, and the reputation of the authors. Finally, we chose to base our systematic review on a total of 22 papers.

3 Literature Review

In this section, we will present an overview of the most distinguished studies in this field. We will highlight the studies that attempted to explicitly or implicitly answer our research questions. This section is divided into four sub-sections, each sub-section will address a group of questions within the same category as follows: SM platforms (questions 1–3), knowledge and information (4–6), frameworks or methods (7–11), and case studies (12).

3.1 Social Platforms in KM

In this subsection, we will be attempting to answer the first three research questions previously addressed in the research method section concerning different social media platforms by highlighting the most used and discussed platforms in the field of knowledge management in a timely manner. We will present the most platforms being discussed by authors and researchers, compare how are they being utilized, and show some statistical graphs and overviews.

In 2006, Gabriela Avram introduced an exciting classification for what she has called Social Software [2]. During that time, weblogs or blogs were popular social platforms. A blog is a personal website in which the site's owner would post a subject allowing the blog visitors to interact with the subject by adding their comments. Also, Wikis began to rise and became a very significant source of information. Thus, the author categorized Social Software into three main categories: weblogs, wikis, and social network services (what is being known today as social media platforms). However, during the year 2006, very few of today's significant platforms were actually in use, such as Wikipedia and LinkedIn, whereas many others have been long gone like: Flickr and Friendster.

In 2010, Plotnick and White referred to social media as collaborative services based on Web 2.0 technologies [9]. The phrase "Web 2.0" was first appeared in 2004 by O'Reilly Media [10]. It encompasses all web services that rely primarily on users' collaborations, interactions, and communications to contribute to particular information space. Plotnick and White included all of the current and recent platforms in their so called collaborative services such as: Twitter, Facebook, Myspace, as well as blogs and wikis.

In 2012, Georg von Krogh proposed a strategic research agenda toward social media and knowledge management [11]. He again linked social media with Web 2.0 technologies and online communities, describing them as blogs, wikis, social bookmarking, data mashups, editing platforms, and media sharing. He gave an example of Facebook and YouTube.

A more recent case study by Nanyang Technological University in 2013 has shown that there are more efforts to utilize some of today's popular networks such as Facebook and Twitter [12]. In this paper, the authors categorized social media services into four main categories: microblogging services, social networking services, location-aware mobile services, and corporate discussion forum services. Social networking services being what is referred to today by social media. The authors have given some services examples such as Facebook, Myspace, and Foursquare.

In 2017, two faculty members from the University Technology of Mara UiTM Selangor in Malaysia presented a case study on the effectiveness of social media as a KM tool in the government sector [13]. In their study, they managed to show a statistical graph comparing different social and chatting platforms in terms of active users as shown in Fig. 1.

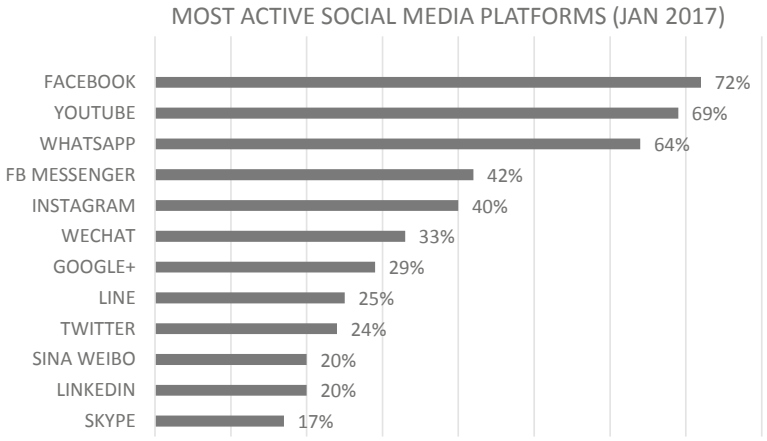


Fig. 1 Most active social media platforms as of Jan 2017 [13]

3.2 The Knowledge in Social Media

This sub-section will focus on the quality of knowledge found in social media platforms. We will review the answers to questions 4–6 from the literature point of view by looking for how the information presented in these platforms is being regarded in the field of KM and has been any effort toward verifying the originality of these information. Gabriela Avram has again classified social software platforms into an exciting classification, but this time based on their content [2]. Some are general networking sites that contain discussions, invitations, and interactions in various topics similar to today's most popular platforms like Twitter, Snapchat, or Facebook. Others are more business-oriented like LinkedIn or Ryze. Also, some platforms are location-based communications and interactions like MeetUp. These platforms build upon invitations, virtual friends, and knowledge sharing containing a vast range of information about any given subject or topic. Most, if not all, platforms do share a common thread, which is providing ways of expressions for shared information or distributed knowledge using sharing techniques like bookmarking, image tagging or sharing, and search engines [14]. Those platforms do play an essential role in supporting knowledge transfer by providing an easy opportunity for everyone to contribute to this cumbersome process. Besides, social platforms have distinct features that motivate users to engage and participate in what so-called a knowledge-sharing environment [12]. The openness of these services makes them a very fertile area for up-to-date information and keeps the knowledge very active and dynamic comparing to traditional passive sources of information such as books or newspapers. Social network services support interconnectedness between customers and organizations through constructive dialogues and live conversations. These ways of interactions serve as a win–win situation for

both sides and produce a massive accumulation of shared knowledge about the organization, products, and customers that significantly aid both sides.

Wiki is one of the most excellent examples of how social platforms can be a significant source of information. It enables people to easily collaborate in order to create specific content [11]. It is not only a tool that allows new content to appear but also allows those contributions and their contributors to be evaluated, which improves the quality of the content effectively and efficiently through rating, editing, and frequent contributions. In general, the knowledge accessible in social networks is flexible to certain circumstances, based on need, and configured to fulfill a specific purpose. Blogs, on the other hand, have also proven to be a highly useful KM tool [15]. Blogs attain their significant powerfulness by spreading information, encouraging comments, and through entries classification and archiving. Blogs easily surpass older technologies such as email or forums especially in business organizations. Despite being one of the most popular social media platforms for a long time, studies are still very narrow when it comes to Facebook. A relatively recent study has attempted to examine the use of Facebook in the academic field specifically in public libraries [16]. The significant amount of posts, sharing, and ratings (likes) have suggested that Facebook is indeed an auspicious tool for Knowledge Management.

3.3 Methods and Frameworks to Manage the Knowledge in SM

In this part, we will capture the most distinguished attempts among researchers to effectively seize, evaluate, combine, and organize the knowledge available in various social channels. In this segment, we will be looking into different contributions that are related to questions 7–11. Avram [2] adopted in her paper Despres' and Chauvel's [17] framework for categorizing KM using five activities:

- Scan/Map.
- Acquire/Capture/Create.
- Package/Codification/Representation/Sorting.
- Apply/Share/Transfer.
- Reuse/Innovate/Evolve/Transform.

In the first activity, she suggested scanning the platforms simply by reading, monitoring, and subscribing to specific searches. Whereas acquiring or capturing the knowledge in those channels can be accomplished by adopting an open editing facility such as wiki or using blogs to expose new ideas to the public and acquire them to share their ideas. Also, manually browsing these platforms and bookmarking interesting posts or pages can serve as a handy acquisition or capturing tool which can be associated with the third activity of packaging and codification as well. Sharing, of course, is the core basis of these social platforms. Contents can be

easily shared between users, and a user can stay up-to-date with a particular development by merely subscribing to its feeds. The University of Florence in Italy went a bit more technical on how social software can support the management of a person's knowledge [14]. They claim that this can be achieved by the process of making meaning of socially shared and co-created content. The author hypothesizes that users have to obtain a set of skills, of which they called it Personal Knowledge Management skills (PKM). However, PKM is not enough on its own. It needs to be associated with an Instructional Design Method (ID). Table 1 summarizes the PKM skills needed by users to be able to benefit from the knowledge in the social platforms.

The study heavily emphasizes on the importance of users' participation in this process of knowledge circulation through the use of various social platforms and their built-in features. It attempts to highlight the significance of active users, contributions, and other personal skills.

To acquire such skills, the study stresses the necessity to apply the (ID) process which consists of four different phases: access, recognize, construct, and develop. Those phases can be best summarized as follows:

13. **Phase 1**: access to online resources and available social media spaces. This phase uses the principle of activation, which is used to for pre-knowledge development.

Table 1 Personal knowledge management skills (PKM) [14]

Create	Organize	Share
• **Editing**: utilizing technology to create digital content for various platforms	• **Searching and finding**: effectively using online search engines, filtering tools, refiners, etc.	• **Relating with people**: initiating connections, effective communication using the latest channels, understanding others, using multi-languages
• **Integrating**: to post-process auto abstracts, digital explanations, recordings, etc.	• **Selective retrieving**: excess and redundant information management	• **Contacts management**: bookmarking profiles and forms of social contacts
• **Correlating**: linking the dots, creating graphs and diagrams, etc.	• **Storing**: archiving to ensure availability and facilitate accessibility	• **Collaboration:** sharing assignments and tasks, working toward common goals
• **Manage security:** privacy protection, IP rights, and digital rights management (DRM)	• **Categorizing/classifying**: differentiating and separating bits and pieces using descriptors	• **Knowledge exchanges**: being short and concise, being able to take turns, concentrating on a particular topic, etc.
	• **Evaluating**: extracting meaningful data, quality assessment, and relating different information	• **Publications**: presenting knowledge and information, using relevant publication media (blogs, wikis, online archives, etc.)

14. **Phase 2**: recognize the objectives of certain information and initiate a knowledge exchange mechanism using the demonstration principle for knowledge acquisition.
15. **Phase 3**: constructing the knowledge and initiating collaborations. In this phase, the person starts applying the knowledge gained and receiving factual information in return.
16. **Phase 4**: Finally, the user starts integrating the acquired knowledge and generating in different communities and groups of his/her interest

Chua and Banerjee proposed a slightly different framework to manage Starbucks customers' knowledge in very current social media channels [12]. They combine a qualitative case study with netnography (a qualitative technique that pulls data from computer-based communication medium) to examine the use of the social network for Customer Knowledge Management (CKM). The framework integrates customer knowledge management techniques with social platforms. The study consists of two main phases: data collection and text analyzing. In the first phase, they performed a search based on pre-identified keywords through specialized search engines like Factiva, LexisNexis, and other publications databases. The final data were analyzed using TexSTAT to generate lists of frequent words to recognize the commonly used social platforms. Netnography was then used to gather data from the commonly used social services. In the second phase, a final text analyzing was performed on the collected data using unitizing and categorizing [18], 19].

The significance of social media relies heavily on the natural and quick access to information, expertise, databases, and other resources [11]. However, the quality of information, knowledge, and data has always been and will be a matter of concern. The lack of formal validation and proper quality control found in conventional knowledge management is a huge downside. What knowledge should be obtained by social media, for what use, and how-to asses its quality? Creating monitored repositories with dedicated "knowledge managers" seeking quality improvement by carefully editing the collected data may contribute to social media data quality improvement in the long run [20]. Members involved in social media data collection in a specific firm should possess high-quality standards when collecting online data. The quality of their knowledge will significantly impact their knowledge management practices. Organizations might fail to embrace the practice of KM due to their inability to aligning business and KM strategies [21]. Previously, the integration of IT to business faced similar issues to what KM is currently facing, as both of them turned from playing a supporting role into playing a vital role in shaping the business strategy of the company.

3.4 Case Studies

In this part, we will highlight some of the successful case studies that have proven the effectiveness of knowledge management in social media. In 2013, a comparative

study by the University of Hong Kong had examined the possibility of applying the basic knowledge management processes of capture, share, and apply among university students [15]. The study focused on two of the major platforms, back then, which are Blogs and Facebook. They performed a comparative analysis to determine which platform can be more facilitating when it comes to college students. The study concluded that although Facebook has a friendlier user interface and generally attract students to participate more effectively in the knowledge sharing process, both platforms are considered to be of significant aid to students in constructing their knowledge base and utilizing their information most efficiently. Another relatively recent study that is more business-oriented than the previous one by Chua and Banerjee has focused on customer knowledge management (CKM) by analyzing the use of different social platforms in a business organization that depends on bricks-and-mortar traditional model, in this case, Starbucks [12]. Starbucks has made a significant effort in the deployment of multiple social media platforms to transform its vast customer base from being primary passive customers to actual contributors and innovators. Moreover, with online promotions and marketing strategies using third-party social platforms as well as their platform "MyStarbucksIdea," Starbucks has managed to encourage their customers to engage in social media. The study concluded that social media is, in fact, a significantly powerful tool not only for online businesses but also for traditional retail ones.

The following research was conducted to study the roles, implications, and effectiveness of the social media technologies that were used by the U.S. government to obtain, share, reuse, apply and maintain knowledge during their response to the 2010 Haiti Earthquake [22]. Technologies such as Wikis and collaborative workspaces were used as part of the knowledge management systems for the mission. It was the first time the U.S. government has decided to rely heavily on social media to manage knowledge and action between the cooperating agencies, including the U.S. Agency for International Development (USAID), the U.S. State Department, and the U.S. Armed Forces.

For example, Microsoft SharePoint was used which enhanced knowledge sharing capabilities due to its provided options of creating web pages by any employee and the ability to tag each contributors' name to their contributions along with their contact details so if anyone had additional information regarding the same issue; they would know whom to contact. Microsoft PowerPoint was used as well to brief the senior leaders of the next 12 h working cycle with the gathered information from each functioning area from the previous working cycle, which created formal and informal tasks for the next cycle. While SharePoint was focused on sharing information within staff, Wikis were used for the holistic relief mission. The benefits of using social media tools in knowledge management were the growing rate of knowledge reuse among staff. Previously, knowledge between staff was only shared during formal briefings where conversations were limited, and the know-how of acquired knowledge of each function was ambiguous to other employees. This problem was solved by applying social media tools as it increased the visibility among employees to see how their colleagues were managing knowledge.

Another benefit is eliminating the reliance on formal liaison structures in terms of personnel and systems for communications among staff. Earlier, the problem was not that employees were prohibited from accessing knowledge of other departments; instead, it was to know where, what, who, and how to access the needed knowledge. However, as mentioned earlier, using shared platforms such as Wikis resolves this problem as it gives access to everyone with references of whom to contact regarding each contribution. Finally, one of the most important benefits was that the reuse of shared knowledge has significantly eradicated unnecessary repetition of effort. Certain challenges were faced due to using social media tools for knowledge management purposes during the disaster response mission. One of the most critical issues during managing emergency response was the need to check the accuracy of the gathered information continually and to validate it which was very challenging due to the massive amount of received data. This could have led to other challenges such as information overload, which could threaten the ability to manage and use the acquired knowledge properly. This is why the organizations need to set processes or system controls to monitor and assess the acquired and shared knowledge. The above case study has demonstrated the potential of Social Media technologies in disaster response. When the technologies are optimized, they produce faster decision cycles and more comprehensive knowledge resources. Behringer and Sassenberg conducted an organizational based study that focused primarily on employees' intentions to utilize local or online social media platforms for knowledge management purposes [23]. The study was conducted on 315 employees within a single organization. They concluded that an individual's understanding of the significance of the knowledge exchange concept and what to expect from it is what drives his/her intention to use social spaces to exchange knowledge. Social media is more or less a technological tool that can be utilized by individuals who are willing to contribute their knowledge and are fully aware of the importance of the sharing concept.

4 Results

Using the framework outlined in Sect. 2, and based on the contributions highlighted in Sect. 3, this section will address the research questions by summarizing our findings in the literature.

What platforms are earnestly being considered by KM researchers, and how up-to-date are they considering today's channels and spaces? It can be easily noticed that social media platforms significantly change over time. Blogs and forums were massively vibrant in the early 2000s studies; however, they began to disappear in the more recent studies. More modern platforms such as Facebook and Twitter have taken the attention of researchers due to their widespread among internet users. However, we have noticed the absence of some of the viral media nowadays, such as Snapchat, TikTok, and SoundCloud. Moreover, very little or none of the studies have focused on platforms that rely on audio/video content only. These platforms might be the richest when it comes to actual genuine knowledge.

What platforms do hold or contain original, verified, and genuine knowledge within their massive content? Most authors tended to focus on Wikis for original and genuine content overlooking its open-source origin and the fact that any individual or user can edit or write an entire article without any constrains. Some Wikis do have certified editors and reviewers such as Wikipedia, while others heavily rely on the user to verify the received information. Facebook is another platform that has gained the attention of researchers for a social space that holds a considerable amount of knowledge. Unlike Wikis, Facebook does not hire any editors or reviewers to trace the originality of what is written in its pages. However, recently, there have been some decent attempts by social media corporates to target and limit fake news including Facebook and Twitter.

Are there any proposed tools, frameworks, or methods to effectively capture the knowledge in various social platforms, organize and create a reliable source of information out of it? While many authors suggested to traditionally and manually gather related data and information simply by scanning, skimming, bookmarking, filtering, etc. some proposed frameworks to facilitate this manual process. Almost all of these frameworks are still heavily dependent on an individual's set of personal skills including knowledge acquisition, personal evaluation, cognitive categorization, and the person's realization of the importance of knowledge contribution. However, we have noticed that there is a lack of automated tools or computer algorithms to spontaneously capture relevant information based on the user's point of interest while many of these social media companies do have artificial intelligence algorithms embedded within their sites to ease the user experience and for other profitable purposes, these tools are not being utilized for knowledge management purposes yet.

Are there any actual successful case studies that demonstrate the significant role of social media in knowledge management? Social media platforms have indeed proven to be one of the most significant sources of knowledge nowadays. In Sect. 3.4, we underlined some of the positive contributions to support this statement. While organizations or even researchers may vary in the way they handle or manage this source of knowledge, all of them have agreed that if proper approaches were implemented to capture relevant data on different social platforms, the results should be promising since these channels or spaces contain a very untraditional type of information. A type that cannot be found elsewhere, because they heavily rely on users' experiences and inputs which, again, cannot be found in traditional knowledge means.

5 Conclusion

This systematic review has highlighted some of the distinguished contributions in the field of managing knowledge in social media platforms. The study addressed specific research questions underlying significant concerns related to knowledge and information verification, tools or frameworks that can assist in capturing

relevant information and some successful experiments or initiatives in the industry. While many efforts and attempts have been made to manage today's most active platforms, most authors and researchers focused on the written materials such as Blogs, Wikis, and Facebook tended to overlook the significance of other types of media such as YouTube, SoundCloud, and SnapChat. Moreover, many of the arising platforms have been neglected. Some outstanding approaches and frameworks have been proposed to effectively capture the knowledge in SM, verify it, and turn it into a reliable knowledge reservoir. These frameworks were supported by experimental case studies that concluded with promising results. However, almost all approaches were limited and bounded within traditional KM processes that suggest manually handle the enormous amount of information presented in these media. There were no apparent attempts for technological or intelligent tools that can automatically extract, verify, and manage the knowledge in different social channels. To conclude, we believe that while traditional KM processes and primary individuals backgrounds can be relied on to a certain extent to extract genuine knowledge from Blogs, Wikis, or other written medias, a numerous amount of data and information is being shared and exchanged on hourly bases that are being wasted and in desperate need of an actual intelligent computer software to extract it and turn it into a knowledge repository.

References

1. Bjørnson FO, Dingsøyr T (2008) Knowledge management in software engineering: a systematic review of studied concepts, findings and research methods used. Inf Softw Technol 50:1055–1068. https://doi.org/10.1016/j.infsof.2008.03.006
2. Avram G (2006) At the crossroads of knowledge management and social software. Electron J Knowl Manag 4(1):1–10
3. Davenport TH, Prusak L (1998) Working knowledge: how organizations manage what they know. Harvard Business School Press
4. Stata R (1996) Organizational learning: the key to management innovation. In: Starkey K (ed) How organizations learn. Thomson Business Press, London, pp 316–334
5. Kaplan AM, Haenlein M (2010) Users of the world unite! The challenges and opportunities of social media. Bus Horiz 53(1):59–68
6. Boyd D (2005) The significance of social software
7. Kane G (2017) The evolutionary implications of social media for organizational knowledge management. Inf Organ 27. https://doi.org/10.1016/j.infoandorg.2017.01.001
8. Kitchenham BA (2004) Procedures for performing systematic reviews. Technical report TR/SE-0401. Keele University
9. Plotnick C, White L (2010) A social media tsunami: the approaching wave. Int J Inf Syst Crisis Response Manage 2(1):i–iv
10. O'Reilly T (2005) What is Web 2.0: design patterns and business models for the next generation of software. O'Reilly Media. Available at: www.oreillynet.com/lpt/a/6228/. Accessed 10 Mar 2009
11. Krogh G (2012) How does social software change knowledge management? Toward a strategic research agenda. J Strateg Inf Syst 21:154–164. https://doi.org/10.1016/j.jsis.2012.04.003

12. Chua A, Banerjee S (2013) Customer knowledge management via social media: the case of Starbucks. J Knowl Manage 17. https://doi.org/10.1108/13673271311315196
13. Dahri A, Yunus A (2018) The effectiveness of social media as knowledge management sharing tool in government agency: a case study. Int J Acad Res Bus Soc Sci 7 https://doi.org/10.6007/IJARBSS/v7-i12/3750
14. Pettenati MC, Cigognini E, Mangione J, Guerin E (2007) Using social software for personal knowledge management in formal online learning. Turkish Online J Distance Education 8
15. Chan RC, Chu SK, Lee CW, Chan BK, Leung CK (2013) Knowledge management using social media: a comparative study between blogs and Facebook. Proc Am Soc Info Sci Tech 50:1–9. https://doi.org/10.1002/meet.14505001069
16. Chu SKW, Du H (2013) Social networking tools for academic libraries. J Librariansh Inf Sci 45(1):64–75
17. Despres C, Chauvel D (1999) Knowledge management(s). J Knowl Manage MCB University Press 3(2):110–120
18. Stathopoulos I, Harrison K (2003) Study at master's level by practising physiotherapists. Physiotherapy 89(3):158–169
19. Chua AYK (2011) How web 2.0 supports customer relationship management in Amazon. Int J Electron Customer Relat Manage 5(3–4):288–304
20. von Krogh G, Haefliger S, Spaeth S, Wallin W (2012) Carrots and rainbows: motivation and social practice in open source software development. MIS Q 36:649–676
21. Phosaard S, Wiriyapinit M (2011) Knowledge management via facebook: building a framework for knowledge management on a social network by aligning business, IT and knowledge management. In: Proceedings of the world congress on engineering 2011, WCE 2011, vol 3, pp 1855–1860
22. Yates D, Paquette S (2011) Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. Int J Inf Manage 31:6–13. https://doi.org/10.1016/j.ijinfomgt.2010.10.001
23. Behringer N, Sassenberg K (2015) Introducing social media for knowledge management: determinants of employees' intentions to adopt new tools. https://doi.org/10.15496/publikation-7724

Overviews of Uncertainty: Concepts, Categories and Coping Strategies in Decision Making

Aisyah Ibrahim, Tuty Asmawaty Abdul Kadir, Hamdan Daniyal, and Adzhar Kamaludin

Abstract Uncertainty is not a new issue in decision making. A wrong decision made of a lack of certainty is every decision maker's nightmare. Although uncertainty is broadly discussed in academic literature, uncertainty has been defined in many definitions, terms and conceptions. Some of the terms often overlap with similar terms from different perspectives. In the light of this, this paper takes the opportunity to explain uncertainty under 3 C's perspectives; the Concept, the Categories and the Coping Strategies. An overview of uncertainty concepts provides the basis for understanding how it is defined from a particular perspective related to decision support for decision making. The categories of uncertainty are explained based on their characterisation to provide a better understanding. Since there is no fixed way to treat uncertainty, this study takes the opportunity to compile the possible coping strategies in handling uncertainty. As an outcome, the contributions of this study come in two ways. First, this paper provides insights on uncertainty concepts and uncertainty categories that are worthy of being reflected in decision support for decision-making research domains. Second, this paper provides solutions and ideas for potential mix-and-match coping strategies that might be helpful in dealing with uncertainty in decision-making.

Keywords Uncertainty concept · Uncertainty categories · Coping strategies · Decision making

A. Ibrahim (✉) · T. A. A. Kadir · A. Kamaludin
Faculty of Computing, Universiti Malaysia Pahang, Pekan, Pahang, Malaysia

H. Daniyal
Faculty of Electrical and Electronic Engineering Technology, Universiti Malaysia Pahang, Pekan, Pahang, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_61

1 Introduction

Uncertainty is a critical issue in decision making [1–3]. Decision-makers often find it hard to decide due to uncertainty. Most of the time, decision-makers are afraid to make bad decisions. Often, bad decisions can lead to adverse consequences. That is why considering uncertainty in decision making can reduce the chances of making wrong decisions [4]. Although understanding uncertainty may take time, it provides significant advantages. First, it helps decision-makers, analysts, and modellers recognise the uncertainty types they are dealing with [5, 6]. Second, it helps decision-makers, analysts, or modellers to frame and reasoning the uncertainty in the proper context before selecting decision supports [2, 5, 7]. Third, it helps to determine which decision support treatment is suitable to be applied to solve the problem or to reach a decision [2, 7–9].

To fully incorporate uncertainty into decision support models or decision-making processes, defining what is meant by the term is vital. Unfortunately, uncertainty definitions are often framed in many definitions. Moreover, some definitions contradict one another. Therefore, this paper takes the opportunity to explain Uncertainties under '3C perspectives': Concept, Categories and Coping strategies. Section 2 explains the methodology used in developing this paper. Section 3 covers the uncertainty concepts and definitions. Section 4 discusses the characterisation of the uncertainty categories. Section 5 highlights the potential strategy to cope with uncertainty, and lastly, Sect. 6 provides the closing remarks.

2 Methodology

Within this process, Google Scholar, Science Direct, SpringerLink, Taylor and Francis and Scopus databases were used to search for publications. Keywords, such as 'uncertainty meaning', 'definition of uncertainty', 'uncertainty concept', 'uncertainty in decision making', 'uncertainty in decision support', 'uncertainty and decision making', 'uncertainty and decision support', 'decision making under uncertainty', 'decision making under uncertain situation', 'uncertainty types', 'uncertainty categories', 'uncertainty groups', 'uncertainty classifications', 'dealing with uncertainty', 'managing uncertainty', 'coping with uncertainty', and 'addressing uncertainty' were used to guide the searching engines. Only papers associated with the domains of interest were considered for selection. All collected papers were reviewed. Content analysis was used to analyse the literature using ATLAS.ti 9.

3 Uncertainty Concepts and Definitions

Uncertainty is defined in many denotations. Some use the word 'risk', 'ambiguity', 'turbulence', 'equivocality', and 'conflict' to describe uncertainty [8]. The way uncertainty is defined depending on how uncertainty is viewed [10, 11]. Generally, uncertainty is associated with knowledge limitation [10, 12]. In decision problem context, uncertainty can be referred to as a decision situation that is not known by the decision-maker or as an action that has not yet been decided by the decision-maker [12] or a situation when the absence of precise information occurs [13]. Most classic literature on uncertainty in decision making concentrates on two limitations. One is knowledge limitation regarding the situation [3, 14–18] such as lack of understanding of the nature of uncertainty state, limited knowledge about the past, the current, and the future situation. And the other one is knowledge limitation about the available options [12, 19], such as the inability to select suitable choices, the inability to see the effects of particular choices, and the inability to rank the choices according to preferences.

Even if the problem in the decision situation is understood, the information in hand might be inadequate, inexact, inaccurate, vague, incomplete or hard to fathom. Surprisingly, having enough data does not guarantee that uncertainty can be eradicated. The truth is, even with ample information or too much available data/information, uncertainty can exist [15]. New information may suddenly appear while decision-making is still in progress. Sometimes, it can reduce uncertainties and provide a better understanding of problematic situations [15, 20]. However, in some other times, this may increase uncertainty where further confusion continues to occur.

In an uncertain situation, the degree of uncertainty will continue to arise if the decision-maker does not know about each option's positive and negative effects and impacts. Of course, sometimes knowing alone, is not enough. For example, knowing that the COVID-19 outbreak pandemic is a global health crisis does not mean the decision-maker/policy maker knows how it will affect the whole country. He/She will always wonder what could happen next and what he/she should do next. (e.g., what actions should be taken next? What are the effects and consequences of actions taken? How effective the policy introduced? Is there any other alternative?) Unfortunately, most of the time, the decision-maker must rely on the knowledge he/she has, to make a decision [12]. Here is where decision support comes to play. But before that, it might be wise to address the uncertainty first, then the coping strategies. The following section discusses the characteristics of the uncertainties and their categories.

4 Categories of Uncertainties

Understanding categories of uncertainty can help decision-makers, modellers, and decision support analysts to recognise the uncertainties they are dealing with. Based on the literature, uncertainty is categorised into three dimensions: by source, by nature, and by level [15, 21–24]. These categories are shown in Fig. 1.

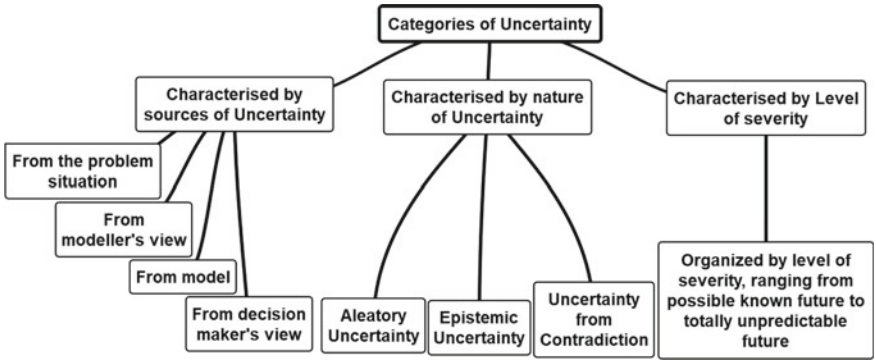


Fig. 1 Categories of uncertainties

It might be good practice for a modeller to list the uncertainties based on their source, their nature of uncertainty, and their severity level before model development starts [9]. Some uncertainties can be viewed from all three dimensions, while others might be beneficial to be perceived based on only one or two dimensions. Doing this helps to identify which uncertainties are critical, which ones need to be prioritised, and which ones should be put aside and ignored. Moreover, categories of uncertainty have a direct influence on the coping strategy chosen. Indirectly, this will also impact the methodology selection in the overall decision process.

4.1 The Ones that Characterised by Sources of Uncertainties

This type of uncertainty are very case-specific [25]. In this category, uncertainty is identified from the source that causes the uncertainty [6, 15]. There are four primary sources of uncertainty. Each is explained as follow:

- From the problem situation—This type of uncertainty arises from the situation. It is understood that different decision situation will involve various sources of uncertainty. For examples: Uncertainty due to variability, goal uncertainty [26], political uncertainty, context uncertainty [15], time scales uncertainty [11, 27]
- From analyst/modellers' view—This type of uncertainty mainly regarding incorporating uncertainty into the model. For examples: technical uncertainties, methodological uncertainties and epistemological uncertainties [6].
- From model—Mostly, this involves the uncertainty that manifests itself within the model [15, 22]. These can stem from context and framing uncertainty, input uncertainty, model structure uncertainty, parameter uncertainty, model technical uncertainty, uncertainty about model validity, and system boundary uncertainty [28].

- From the decision-maker's view—This type of uncertainty comes from the problem owner's preferences, needs, goals and desires. For examples: uncertainty on agenda-setting, cost and benefits of alternative options, and goals and preferences [6].

4.2 The Ones that Characterised by Nature of Uncertainties

This category is related to the nature of knowledge. There are two types of uncertainty associated with the nature of uncertainty: Aleatory Uncertainty and Epistemic Uncertainty [9, 21, 23, 24]. Aleatory or stochastic Uncertainty is an ontic Uncertainty that usually associated with inherent variability or randomness. This type of uncertainty is irreducible. Unlike the former, Epistemic Uncertainty can happen due to a lack of knowledge about the situation or due to ambiguity [29]. Ambiguity in this context means that there exist multiple frames of reference about given phenomena. Epistemic Uncertainty can be reduced by further research and empirical investigation or new theories based on experimental observation [30]. According to authors [28], there is another type of nature uncertainty called 'Contradiction' in the literature. Contradiction Uncertainty is an emergent category from ambiguity situation, for "knowing too differently". This situation happens when 'two or more equally plausible interpretation possibilities exist' [9, 15]. For examples, the ones associated with knowledge against ignorance, facts against values, and science against the policy. In this situation, criteria of choice are critical as this will determine the decision-makers decision course.

4.3 The Ones that Characterised by the Level of Severity of Uncertainties

In this category, the uncertainty is characterised based on the degrees of severity in a one-way continuum. The measure is from a low level of uncertainty to an extremely high level of uncertainty. Table 1 is showing several types of level of seriousness proposed by researchers in the literature. By understanding the level of uncertainty, it helps to assist:

- on how to propagate uncertainty through a model, or whether it cannot be, and must simply be acknowledged as the present [28].
- to determine suitable treatments. As long as possible futures can be known, predictions, projections and forecasts are still practical methods to employ [33]. As the unknown is reached, it can be explored and speculated. To some extends, it is wise to ignore it as it is out of the boundary knowledge [34].

Table 1 Examples of uncertainty level based on degrees of severity in few selected literatures

Source of literatures	Uncertainty level used in their work
Kwakkel et al. [31]	1. Shallow uncertainty 2. Medium uncertainty 3. Deep uncertainty 4. Recognised ignorance
Walker et al. [3]	1. A clear enough future (with sensitivity) 2. Alternate futures with probabilities 3. Alternate futures with rankings 4. A multiplicity of plausible futures (unranked) 5. Total unknown
Koul et al. [17]	1. High level of uncertainty 2. Uncertainty that was caused by multiple plausible options 3. Uncertainty with the basis of probability 4. Uncertainty that people know 5. Uncertainty that people know that they do not know
Thissen et al. [32], Pruyt and Kwakkel [20]	1. Recognised uncertainty 2. Shallow uncertainty 3. Medium uncertainty 4. Deep uncertainty 5. Recognised ignorance
Walker et al. [23]	1. Little uncertainty 2. Statistical uncertainty 3. Scenario uncertainty 4. Deep uncertainty

- on how to distinguish between bounded uncertainty (where all possible outcomes are believed to be 'known') and unbounded uncertainty (where some or all possible outcomes are not known) [25]. Unbounded uncertainty can happen when everything is completely dark and the unknown is too deep. This type of Uncertainty is known as 'Deep uncertainty' or "Great Uncertainty" [12]. Deep Uncertainty happens 'where analysts do not know, or the parties to a decision they cannot agree on' [35]. If a situation is too complicated or too deeply unknown, this can be classified as a wicked problem [36–38]. In this type of uncertainty, information missing can be a wide range of aspects of the decision. This includes what issues will be settled in the decision; the appropriate time for decision making, what aspects of the decision outcome should be included in the evaluation of the options, and perhaps many more [12].

5 Strategies to Cope with Uncertainty

Once the uncertainty category has been determined, the next step is to select a suitable strategy to cope with the uncertainty. Since uncertainties do not necessarily manifest in similar forms and levels, many strategies and tactics have been explored

and applied to deal with uncertainty [39]. To our best of knowledge, there are no specific set of frameworks to deal with all kind of problem case. A single problem case may be treated more than one treatment. So that is why one particular problem case may have many different solutions. The exact selection will vary from case to case. Considering this factor in mind, this paper only reviews the coping strategies in general instead of narrowing down specific uncertainty treatments for specific uncertainty types. This is because the coping strategy can be employed in combinations based on how the problem is framed, the types of uncertainty and the method chosen.

There are the seven coping strategies that have been identified in our analysis: (1) acknowledging uncertainty, (2) reducing uncertainty, (3) suppressing uncertainty, (4) framing uncertainty, (5) methodology selection, (6) stakeholders involvement and (7) computerised decision supports. Each is explained as follows:

- (1) Acknowledging Uncertainty—This approach requires the decision-maker to recognise the source and the types of uncertainty. Usually, these tactics are applied when reducing uncertainty is unavoidable, or not feasible or too costly. In this case, either one of these two valuable strategies can be chosen. One is by taking account in the choice selection, and the other one is by preparing to avoid or confront risks [8].
- (2) Reducing Uncertainty—This approach aims to lower down the uncertainty or to remove it altogether. Reducing uncertainty can be done through several ways: (i) experimental observation [30], (ii) using extrapolation from available information using statistical methods to predict future, (iii) using extrapolation from available information using assumption-based reasoning, (iv) using mental simulation (v) through scenario building (vi) through information processing [8].
- (3) Suppressing Uncertainty—In this approach, two common tactics can be used. One is called tactics of denial, and the other one is called tactics of rationalisation. These approaches help decision-maker from falling into paralysis, especially when they cannot cope with their uncertainty by reduction approach or acknowledgement approach [8].
- (4) Framing Uncertainty—The way uncertainty is framed within a situation can (1) influence the direction of the overall managing process [7], (2) influence how the conclusion(s) is reached [2, 40], and (3) influence how the decisions are made. [2, 41]. A-frame can be perceived as an "unconscious structure for thinking about things and their relations" [41]. This includes "semantic roles, relations between roles, and relations to other frames" [40]. In research, framing is not a subconscious act, but rather a sense-making medium that provides an interpretation of the reality of the situation, which allows reasoning within the situation context [2]. Cognitively, framings provide knowledge structures [7]. Indirectly, slowly, framing can be considered as part of building scientific arguments. Among practical tactics for framing are (1) through interactional approach [7], (2) defining the scope of a claim, (3) defining depth of analysis, (4) setting in the level of belief which allows reasoning within the context situation [2] and perhaps many more.

- (5) Methodology Selection—The author [9] proposed that uncertainty should be tackled through a methodological approach. This can be done in multiple ways. The suggested strategies are:

  a. methodologies to suit the modelling/development process and level of ambition—In this approach, uncertainty is managed differently throughout the modelling/development process stages. For examples:

     - at the early stage of modelling: the uncertainties are identified and organised based on their categories using an uncertainty matrix.
     - In the middle of the modelling process: some uncertainties will be incorporated into the model and will be quantified with the best estimations and assumptions. Some will be ignored since nothing can be done about them.
     - at validation stage: uncertainty analysis is carried out in calibration and validation activity.

  b. methodologies based on sources and type of uncertainty—In this approach, the methodology selection depends on the uncertainty sources and types of Uncertainty involved. For example, the use of Artificial Intelligent methods in searching for information and processing techniques in 'Scenario Discovery' [27, 39, 42] is proven helpful in decision support under poor data situation. Scenario discovery can provide alternatives solutions, especially if the decision-maker is looking for any. In different circumstances, such as during the modelling work, seeking general method like expert elicitations can be the best strategy when data is hard to obtain, or no data at all or too few direct empirical data available to infer uncertainty. This is a valuable strategy for quantitative risk analysis to quantify uncertainties. However, not all methods are suitable to treat many uncertainties at once. Some only helpful for a specific purpose only. For example, Monte Carlo analysis is only suitable to deal with a specific type of uncertainty like Statistical Uncertainty.
  c. methodologies to suit the purpose of use—In this approach, the methodology is chosen based on the aims. For example, a modeller should employ effective data collections and information gatherings for sufficient information to reduce Epistemic Uncertainty. Another example is when facing an uncertain future case where the main goal is to find robust actions [33]. There are two types of strategy-based-analytics solution that seems appropriate to employ: the 'bottom-up' strategy and the 'top–bottom' strategy. The former uses analytic approaches to assess the vulnerability of alternative strategies, and the latter uses analytics to select those actions that do well under a wide variety of possible futures uses [32, 33].

- (6) Stakeholders Involvement—Stakeholders involvement, such as participatory modelling, is critical in specific modelling processes such as System Dynamics, Bayesian Network, Fuzzy Cognitive Mappings and Agent-Based Modelling [43]. Communication with stakeholders through dialogues, meetings or

interviews can help identify needs, expectations and categories of uncertainties, for framings, and for problem identifications which can be beneficial when tackling ill-structured, complex problems [9]. Other useful tactics are stakeholder reviews and expert elicitations. By working together, this can help to (1) to promote a better exchange of views, (2) to accumulate facts and information from a broader source of expertise and knowledge, (3) to eliminate mistakes, misunderstandings or incorrect information, and (4) to close gaps and achieve same level consensus among technical experts and stakeholders involved [44–46].

- (7) Computerized Decision Support Systems—A decision support system (DSS) is a computer-based solution that assists and improve decision-making abilities for a single or a group of decision-makers [47]. Computer technology-based solutions have been the underlying foundation for decision support in complex decision-making and problem-solving since the 1960s [47, 48]. This is useful, particularly in semi-structured and unstructured decision problems [49]. Based on a review of 83 DSS research papers dealing with uncertainties by [1], many methods and techniques have become the workhorse of decision supports. These methods are Probabilistic, Imprecise Probabilistic, Set Inclusion, Scenario Developments and Analysis, Linguistic Imprecision, Linguistic Imprecision, Optimisations, Simulations, Rules, Multi-Criteria and Multi Objectives Models.

6 Concluding Remarks

To date, this paper serves as the first manuscript that comprehensively discuss Uncertainty in the 3Cs; Concept, Categories and Coping Strategies. It is understood that it is hard to eliminate uncertainties entirely. While bearing this fact in mind, this does not mean that uncertainty cannot be tamed down and manage appropriately. Tremendous efforts have been invested in understanding and characterising uncertainties, developing frameworks by utilising various methodologies and decision supports. The strategy can be applied either in combinations with a pair of two, three or more rather than just a single approach alone. However, one must be aware that the selection must be made carefully based on the category of uncertainties that ones are dealing with. This is because there is no ‘one-fits-all solution that exists.

Moreover, some uncertainties are worthy of being considered, whereas some are best ignored. Some strategies might tackle various problems and satisfy many uncertainties, but some strategies can only cater to specific usage. Therefore, this paper might be useful for analysts and modellers in grasping how to frame and address the uncertainty appropriately and select the proper approach in decision support for decision-making betterment.

Acknowledgements This study is sponsored under MyPhD-MyBrain15 Scholarship by the Ministry of Higher Education Malaysia (MoHE). In addition to that, this study is also supported under research grant RDU200756, funded by Universiti Malaysia Pahang.

References

1. Dias LC, Henggeler Antunes C, Rios Insua D (2012) Dealing with uncertainty in decision support systems: recent trends (2000–2011). Intell Decis Technol 6(4):245–264
2. Guillaume JHA, Helgeson C, Elsawah S, Jakeman AJ, Kummu M (2017) Toward best practice framing of uncertainty in scientific publications: a review of water resources research abstracts. Water Resour Res 53(8):6744–6762. https://doi.org/10.1002/2017WR020609
3. Walker WE, Lempert RJ, Kwakkel JH (2013) Deep uncertainty. In: Encyclopedia of operations research and management science. Springer, pp 395–402
4. Byrne KA, Peters C, Willis HC, Phan D, Cornwall A, Worthy DA (2020) Acute stress enhances tolerance of uncertainty during decision-making. Cognition 205:104448. https://doi.org/10.1016/j.cognition.2020.104448
5. Hirsch Hadorn G, Brun G, Soliva CR, Stenke A, Peter T (2015) Decision strategies for policy decisions under uncertainties: the case of mitigation measures addressing methane emissions from ruminants. Environ Sci Policy 52:110–119. https://doi.org/10.1016/j.envsci.2015.05.011
6. Van Asselt MBA (2020) Perspectives on uncertainty and risk. In: Perspectives on uncertainty and risk. Springer, pp 407–417
7. Brugnach M, Dewulf A, Pahl-Wostl C, Taillieu T (2008) Toward a relational concept of uncertainty: about knowing too little, knowing too differently, and accepting not to know. Ecol Soc 13(2):1–14. https://doi.org/10.5751/ES-02616-130230
8. Lipshitz R, Strauss O (1997) Coping with uncertainty: a naturalistic decision making analysis. Organ Behav Hum Decis Process 69(2):149–163. https://doi.org/10.1006/obhd.1997.2679
9. Refsgaard JC, van der Sluijs JP, Højberg AL, Vanrolleghem PA (2007) Uncertainty in the environmental modelling process—a framework and guidance. Environ Model Softw 22 (11):1543–1556. https://doi.org/10.1016/j.envsoft.2007.02.004
10. Anderson EC, Carleton RN, Diefenbach M, Han PKJ (2019) The relationship between uncertainty and affect. Front Psychol 10(November). https://doi.org/10.3389/fpsyg.2019.02504
11. Van 't Klooster SA, Veenman SA (2021) Shaping future perspectives in policy advice under deep, long-term uncertainty. The case of the Dutch Delta Committee. Futures 127:102700. https://doi.org/10.1016/j.futures.2021.102700
12. Hansson SO (2017) Uncertainty and control. Diametros 53:50–59. https://doi.org/10.13153/diam.53.0.1100
13. Molden DC, Higgins Tory E (2004) Categorisation under uncertainty : resolving vagueness and ambiguity with eager versus vigilant strategies. Soc Cogn 22(2):248–277
14. Duncan RB (1972) Characteristics of organisational environments and perceived environmental uncertainty. Adm Sci Q 313–327
15. Walker WE et al (2003) Defining uncertainty: a conceptual basis for uncertainty management in model-based decision support. Integr Assess 4(1):5–17. https://doi.org/10.1076/iaij.4.1.5.16466
16. Bohanec M (2009) Decision making: a computer-science and information-technology viewpoint. Interdiscip Descr Complex Syst 7(2):22–37
17. Koul S, Falebita OA, Akinbami J-FK, Akarakiri JB (2016) System dynamics, uncertainty and hydrocarbon resources modelling: a systematic review. Renew Sustain Energy Rev 59:199–205. https://doi.org/10.1016/j.rser.2015.12.088
18. Pruyt E, Kwakkel JH, Hamarat C (2013) Doing more with models : illustration of a SD approach. In: 31st International conference of the system dynamics society, pp 1–23

19. Conrath DW (1967) Organisational decision making behavior under varying conditions of uncertainty. Manage Sci 13(8):B-487
20. Pruyt E, Kwakkel J (2012) A bright future for system dynamics: from art to computational science and more [Online]. Available: uuid:3f5b22cb-1552-4ef4-88bf-9f6d6deb14c3.
21. Maier HR, Guillaume JHA, van Delden H, Riddell GA, Haasnoot M, Kwakkel JH (2016) An uncertain future, deep uncertainty, scenarios, robustness and adaptation: how do they fit together? Environ Model Softw 81(November):154–164. https://doi.org/10.1016/j.envsoft.2016.03.014
22. Gabbert S, van Ittersum M, Kroeze C, Stalpers S, Ewert F, Alkan Olsson JA (2010) Uncertainty analysis in integrated assessment: the users' perspective. Reg Environ Chang 10 (2):131–143. https://doi.org/10.1007/s10113-009-0100-1
23. Walker W, Marchau V, Bloemen P, Lawrence J, Lempert R, Kwakkel J (2018) Comment on 'from data to decisions: processing information, biases, and beliefs for improved management of natural resources and environments' by Glynn et al. Earth's Futur 6(5):757–761. https://doi.org/10.1002/2017EF000750
24. Curry J (2011) Climate science and the uncertainty monster genealogy of the uncertainty monster. December, pp 1667–1682. https://doi.org/10.1175/BAMS-D-10-3139.1
25. Brown JD (2004) Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. Trans Inst Br Geogr 29(3):367–381. https://doi.org/10.1111/j.0020-2754.2004.00342.x
26. Pruyt E, Hamarat C (2010) The concerted run on the DSB Bank: an exploratory system dynamics approach. In: Proceedings of the 28th international conference of the system dynamics society, January 2010, pp 1–27 [Online]. Available: http://www.systemdynamics.org/conferences/2010/proceed/papers/P1027.pdf
27. Guivarch C, Lempert R, Trutnevyte E (2017) Scenario techniques for energy and environmental research: an overview of recent developments to broaden the capacity to deal with complexity and uncertainty. Environ Model Softw 97:201–210. https://doi.org/10.1016/j.envsoft.2017.07.017
28. Guillaume JHA, Pierce SA, Jakeman AJ (2010) Managing uncertainty in determining sustainable aquifer yield
29. Morales-Torres A, Escuder-Bueno I, Serrano-Lombillo A, Castillo Rodríguez JT (2019) Dealing with epistemic uncertainty in risk-informed decision making for dam safety management. Reliab Eng Syst Saf 191:106562. https://doi.org/10.1016/j.ress.2019.106562
30. Cadena JE, Osorio AF, Torero JL, Reniers G, Lange D (2020) Uncertainty-based decision-making in fire safety: analysing the alternatives. J Loss Prev Process Ind 68:104288. https://doi.org/10.1016/j.jlp.2020.104288
31. Kwakkel J, Walker WE, Marchau VAWJ (2010) Classifying and communicating uncertainties in model-based policy analysis. Int J Technol Policy Manag 10(4):299. https://doi.org/10.1504/IJTPM.2010.036918
32. Thissen W et al (2017) Dealing with uncertainties in fresh water supply: experiences in the Netherlands. Water Resour Manag 31(2):703–725. https://doi.org/10.1007/s11269-015-1198-1
33. van Dorsser C, Taneja P, Walker W, Marchau V (2020) An integrated framework for anticipating the future and dealing with uncertainty in policymaking. Futures 124 (June):102594. https://doi.org/10.1016/j.futures.2020.102594
34. Zurek MB, Henrichs T (2007) Linking scenarios across geographical scales in international environmental assessments. Technol Forecast Soc Change 74(8):1282–1295. https://doi.org/10.1016/j.techfore.2006.11.005
35. Lempert RJ (2003) Shaping the next one hundred years: new methods for quantitative, long-term policy analysis. Rand Corporation
36. Head BW (2014) Evidence, uncertainty, and wicked problems in climate change decision making in Australia. Environ Plan C Gov Policy 32(4):663–679. https://doi.org/10.1068/c1240
37. Ritchey T (2011) Wicked problems–social messes: decision support modelling with morphological analysis, vol 17. Springer Science & Business Media, Berlin

38. Kwakkel J, Walker W, Haasnoot M (2016) Coping with the wickedness of public policy problems: approaches for decision making under deep uncertainty. J Water Resour Plan Manag 142(3):1816001. https://doi.org/10.1061/(ASCE)WR.1943-5452.0000626
39. Gong M et al (2017) Testing the scenario hypothesis: an experimental comparison of scenarios and forecasts for decision support in a complex decision environment. Environ Model Softw 91:135–155. https://doi.org/10.1016/j.envsoft.2017.02.002
40. Lakoff G (2010) Why it matters how we frame the environment. Environ Commun 4(1):70–81. https://doi.org/10.1080/17524030903529749
41. Mitchell M, Roffey-Mitchell T (2018) Uncertainty, risk, and opportunity frames in Australian online media reports of the 2016 great barrier reef mass coral-bleaching event. Cogent Soc Sci 4(1):1–21. https://doi.org/10.1080/23311886.2018.1536317
42. Kwakkel JH, Auping WL, Pruyt E (2013) Dynamic scenario discovery under deep uncertainty: the future of copper. Technol Forecast Soc Change 80(4):789–800. https://doi.org/10.1016/j.techfore.2012.09.012
43. Voinov A, Bousquet F (2010) Modelling with stakeholders. Environ Model Softw 25 (11):1268–1281
44. Towill DR (1993) System dynamics- background, methodology and applications. 1. Background and methodology. Comput Control Eng J 4(5):201–208. https://doi.org/10.1049/cce:19930050
45. Elsawah S et al (2017) An overview of the system dynamics process for integrated modelling of socio-ecological systems: lessons on good modelling practice from five case studies. Environ Model Softw 93:127–145. https://doi.org/10.1016/j.envsoft.2017.03.001
46. Nabavi E, Daniell KA, Najafi H (2017) Boundary matters: the potential of system dynamics to support sustainability? J Clean Prod 140(1):312–323. https://doi.org/10.1016/j.jclepro.2016.03.032
47. Power DJ (2016) Computerized decision support case study research: concepts and suggestions. In: Papathanasiou J, Ploskas N, Linden I (eds) Real-world decision support systems: case studies. Springer International Publishing, Cham, pp 1–13
48. Arnott D, Pervan G (2014) A critical analysis of decision support systems research revisited: the rise of design science. J Inf Technol 29(4):269–293. https://doi.org/10.1057/jit.2014.16
49. Mysiak J, Giupponi C, Rosato P (2005) Towards the development of a decision support system for water resource management. Environ Model Softw 20(2):203–214. https://doi.org/10.1016/j.envsoft.2003.12.019

Implementing a Talent Management System in a Group of Companies—A Case Study from UAE

Abdulla M. Alsharhan, Said A. Salloum, and Khaled Shaalan

Abstract This case study aims to examine the status quo of an information system in the context of talent management (TM), in a leading group of companies that is under the ports and industrial lands sector in the United Arab Emirates. This study identified the major issues encountered by the target organization from the perspective of the employee. It also analyzed prominent state-of-the-art tools that can solve issues related to TM in the targeted organization, and proposed recommended solutions on a TM system implementation. The data obtained was mainly collected using open-ended interviews with the group employees, as well as qualitative information collection by the researcher through participant observation of mergers and acquisitions in the group. In addition, information from previous studies was also leveraged to formulate the case. The findings indicate that the major issues related to TM include career development plan, learning management system, succession plan, data island, and mobile-first. The study concluded that although it might be easier for the organization to extend their current solution to cover TM, a system such as "BambooHR" and "Ceridian Dayforce" may be worth considering to solve the current organizational TM issues. The outcomes of this case study shed light on the current TM challenges faced by a group of companies during mergers and acquisitions, and offer an essential case study for researchers in TM, HR, information systems and knowledge management.

Keywords United Arab Emirates · Talent management system · Group of companies · UAE · Abu Dhabi · Mergers and acquisitions · Human resources · Talent management · IT project · SAAS

A. M. Alsharhan · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_62

1 Introduction

While human resource (HR) management is a well-known and established field, talent management (TM) is a relatively new field that has been growing significantly over the past 10 years. The term appeared in the late 90s when mentioned by McKinsey and their organization in their 1997 paper "The War for Talent". Since then, the field has created enthusiastic interest and attracted many scholars from various fields [1]. The importance of TM comes from the fact that talents are the critical factor for organizational success in the postmodern era. Besides that, TM is considered as a form of investment, where talents become the primary source of innovation and social development [1]. Losing key talents can cause organizational amnesia. In the business context, information systems are used in different business processes to maximize resource utilization and enhance organization performance. HR professionals use information system technologies on a broad scale to improve HR performance [2]. For this study, the target domain sector is a group of companies based in Abu Dhabi in the United Arab Emirates. This group was established in 2006, working in managing and operating a portfolio of industrial zones, commercial plots, and logistics parks. The group is currently contributing a significant percentage of Abu Dhabi's non-oil gross domestic product (GDP) growth. In the last few years, the group has expanded rapidly through acquisitions, merging, and joint ventures. They have acquired many entities in industrial and logistics sectors.

The rapid growth and merge resulted in the group acquiring a pool of talents coming from diverse corporate cultures, including local and non-local staff. The group is facing a situation where they have to keep up their Emiratization rate. On the other hand, the workforce might lose track of their north star. As a result, the company may face talent-related issues such as unutilized resources, demotivated employees, redundant workforce, increased operational expenses, and higher turnover, especially among the company's most talented experts. Information Systems and Knowledge Management (KM) tools can offer many solutions to aid decision-making in HR processes and practices. Although the company has already invested in a comprehensive enterprise resource planning system (ERP), it might not be sufficient to operate a single platform for all integrated applications. Many organizations today are finding the necessity of adding a third-party system specialized in TM [3]. The field of TM has grown significantly in the last 10 years and is gaining the attention of high performing companies. IT service providers are realizing the importance of TM, hence many vendors are now offering TM system solutions. The importance of highlighting this case is to provide an insight into some TM challenges faced by group of companies in the time of mergers and acquisitions. It also provides possible mitigation solutions that can be solved by information systems and KM tools for scholars in HR and TM context.

1.1 Talent Management System in the Targeted Organization

In recent years, the targeted organization has made significant development in its workforce [4–7]. The hard work paid off in 2014 with the organization wining the People Development Award at the International Bulk Journal (IBJ) Awards in 2014 and 2015. The company has also become among the first organizations in the Middle East to receive the "Investor in People (IIP)" Platinum Accreditation, an accreditation that is acquired by only two per cent of "Investor in People" accredited organizations worldwide [8].

Identifying the research gap. There is little known about how the group of companies is managing the information system architecture in the context of TM, especially after merging and acquiring many significant semi-governmental and private organizations. Besides, the field of KM and information system is rapidly evolving. Within the current context of Covid-19, many organizations are rushing into digital transformation without putting people as their priority. It is essential to understand what employees think are the major issues that might be faced by the group. Finally, TM solutions are also adapting to these changes. What used to work five years ago might no longer suit the new generation of workforce. In this case study, the following research questions will be examined:

RQ1 What are the major challenges faced by the targeted organization from employees' perspectives?

RQ2 What are some of the state-of-the-art tools that can solve issues related to TM in the targeted organization?

2 Literature Review

2.1 Talent Management

TM can have different definitions, whether it is in the context of information system or HR. In the context of HR, TM can be defined as the process of identifying prominent roles which differentiate and ensure organizational sustainability, developing a workforce with high potential and performance to fulfill these roles, and developing a HR structure that facilitates filling these vacancies [3]. The Institute for Corporative Productivity defines TM as "an organizational approach to leading people by building culture, engagement, capability, and capacity through integrated talent acquisition, development, and deployment processes that are aligned to business goals" [9]. For this study, TM can be defined as a continuous identification of the core organizational functions and the matching of these functions to high performing individuals with great potential by allowing the organizational structure and culture to attract, engage, develop, motivate, and retain these individuals.

2.2 Talent Management Importance

Talent Management System. The authors of [10] define Talent Management System (TMS) as an integrated software tool that processes four pillars of talent management: (1) recruitment, (2) performance management, (3) learning and development, and (4) compensation management [10]. After considering the two previous definitions, TMS can be defined as a software suite designed to enhance the process of recruiting talents, developing their competencies, tracking their performance, compensating, and retain them.

Talent Management System components. TMS has different understandings which can vary based on organizational cultures and HR professionals' preferences. The author of [11] found that managers perceived TM in different ways. There are no agreed best practices, and it is usually customized to business needs, or according to TMS vendor classification. Some think it is only related to recruiting potential talent for a job [12], while other definitions include managing, evaluating, and compensating talents based on their performance [13]. Some described TMS activities as planning, attracting, selecting, developing, and retaining high performing workforce using integrated suite focuses [14].

The author of [15] proposed that TMS should include at least (1) employee competency assessment, (2) succession planning, (3) training and professional development, and (4) compensation. On the other hand, [2] discussed only three categories: talent acquisition, development, and retention.

Talent Acquisition Technologies. Talent acquisition can be described as the process of attracting and acquiring potential, highly talented and productive individuals. Technological advancement has allowed the talent acquisition process to become much easier in recent years. One example of technological advancement in talent acquisition is the use of live Chabot to filter candidates and reduce load on recruitment professionals.

Talent Development Technologies. Talent development process usually starts after the employee joins the organization, where talents' capacity and capability are aligned with the business need. For example, learning management systems like Blackboard and Moodle are used to manage learning materials in one place, while software like Adobe Captive can be used to design these learning materials with ease. Artificial intelligence solution is also used to evaluate the individual performance and provide periodic feedback. Other technological tools that can increase collaborative working like project management are file sharing and wikis.

Talent Retention Technologies. Retention refers to the period that employees stay within the organization. Technological application can assist in retaining employees using artificial intelligence to predict employee demotivation, which can be avoided before employee resignation. Polls and surveys can also collect and display information on how employees feel toward their workplace as well as areas of improvement [2].

Talent Succession Technologies. Succession planning management refers to the process of identifying and developing talents with high potential to fill in critical

positions in an organization [16]. For example, an application like "Succession Wizard" can simplify the whole process by identifying potential succession gaps and assisting in planning future staffing needs within your organization.

Talent Performance Technologies. Talent performance management refers to the activities of planning, managing, tracking, and evaluating talents. The author of [17] found a significant positive relationship between TM and projects success factor, including the significant role of talent satisfaction and motivation. Talent performance management technologies can allow efforts, achievements, and comments to be captured in one place. SAP, for example, offers managers a streamlined objective setup using wizards and a library of best practices objectives to guide them through the process. HR Management for Microsoft Dynamics 365 offers continuous performance management by allowing progress check at any time.

Talent Engagement Technologies. Talent engagement technologies refer to the tools that can enhance workforce engagement, collaboration, and motivation. An example of motivational tool is gamification which involves elements of game design and experience to strengthen talent collaboration [18]. Another interesting concept that supports cooperation, involvement, and motivation is the use of "working out loud" tools. It is a way to describe people's social involvement and cooperation using storytelling technology [19].

3 Research Methodology

3.1 Identifying the Targeted Organization TMS Issues

The data obtained in this case study was mainly collected using open-ended interviews with the targeted organization employees. Some data was also collected from ex-employees' feedback written on the job listing website "Indeed.ae". In addition to the information from different past studies between 2016 and 2020 to formulate the case, qualitative information was also collected by the researcher through participant observation of mergers and acquisitions in the group. Keywords used in locating the research papers include the targeted organization name AND "Talent Management", "Talent Management System" AND ("implementation" OR "implementing") AND "Knowledge Management". Databases used to retrieve the papers are Crossref, Worldcat and Google Scholar.

3.2 Recommended TMS Solutions

TMS solutions for the targeted organization were evaluated based on available solutions. Keywords used to locate TMS Software are "Talent Management software" OR "Talent Management System software" OR "Talent Management solutions". Scores were determined by several factors. Each factor was collected from

major software reviewing sites that highlight key generic TMS features. Seventeen factors were identified; regular factors were given a one-point rating, while factors related to the current organization issues were given two points.

4 Results

The following is the result of examining the interview answers provided by the targeted organization employees and ex-employees in addition to qualitative observation through participant observation and previous studies to formulate the case.

4.1 RQ 1. What Are the Major Challenges Faced by the Targeted Organization in the Context of Information System?

After analyzing several employees' feedback and testimonials, the challenges related to KM and TM faced by the targeted organization have been summarized in the following points:

1. **Individual performance in the context of organizational performance**. Although the key performance indicator is clear to everyone, it is still unclear to many how their performance is linked to the whole organizational performance, and how their performance compares to peers and other divisions. For example, some individuals may have been given very relaxed objectives, while others may have tough ones.
2. **Learning management system**. The KM team is actively conducting sessions. They are also hosting guests and sharing knowledge once in a while. Unfortunately, the knowledge shared cannot be retrieved or looked for in an accessible platform. The sessions are also not available in different formats or languages (audio, text, video), limiting the learning materials reusability. Besides that, the sessions taking place during specific times limits the distribution of knowledge. The learning impact cannot be measured within the existing setup, and its link to employee development is not clear.
3. **Career development planning**. The Learning & Career Development Team in the examined organization is investing significant efforts to gather and plan learning needs for each division and department. A word document is sent out annually for each department's focal point to specify training needs. Each department is also asked to link the identified training to strategic objectives. However, not only are these practices conducted manually, they also neglect the individual technical and/or behavioral competency skills needed to advance in a potential career path.

4. **Succession planning**. Having a good succession plan in place can motivate employees and increase their retentions, but this may not be the case of many enterprises, including this organization. The group has employees with high expertise in various fields; many of them are highly experienced and occupy high-ranked positions. However, there is no clear process on identifying and developing employees to replace old leaders, managers, or supervisors to prepare for cases of leave, sickness, emergency, retirement, or mortality.
5. **Mobile-first**. As 5G is becoming widely available, more people are turning to their mobile phones instead of their desktop computers to go online. BroadbandSearch.net (2020) indicates that mobile phones made up 52.2% of the combined worldwide traffic, indicating that we may have reached a new normal. Most Ad Ports services are well-functioning when accessed from mobile smart devices. However, the knowledge platforms are not designed to utilize the full experience of mobile usage. Investing in mobile-first can enhance team collaboration via notifications and provide real-time data for top management via push notifications, reminders, and reports. Offering a mobile-first experience can impact both employee and client experience.
6. **Data island**. The company initially used Oracle solutions for ERP but is recently implementing Microsoft Dynamics 365 for their CRM needs, in addition to different software from different vendors such as Zoho helpdesk and MeetingBooster. It is unclear how system integration is working among these different solutions.

4.2 RQ 2. What Are Some of the State-of-the-Art Tools that Can Solve Issues Related to TM in the Targeted Organization?

Many Software as a Service (SaaS) licensing vendors have realized the importance of designing a specific solution for TM, while others are still incorporating some of TM's components in their HR solutions. The following lists are all the TMS features provided by different TM suites available in the market, which have been highlighted by many software monitoring and reviewing sites to evaluate TMS suites such as "Software World", "Software Testing Help", and "GetApp".

Evaluation of alternative solutions. Table 1 shows each identified issue in the targeted organization and evaluation of alternative solutions from a possible TMS component to mitigate or eliminate the identified issue.

After evaluating the alternative solutions, seven essential factors were identified to have impact on the targeted organization. Therefore, they were given two points in the following evaluating table. These essential factors are career development planning, goal setting/tracking, on-boarding, succession planning, learning and development, system integration, and mobile-first (Table 2).

Table 1 Evaluation of alternative solutions

	Major problems	Evaluation of alternative solutions	Possible TMS component mitigation
1	Individual performance in the context of organizational performance	The lack of proper employee onboarding process can lead to a decrease in their engagement in an organization. Besides, the company can enhance team collaboration and motivation by gamification tools. Moreover, working-out-loud techniques also play significant roles in encouraging performance transparency, lifting talent moral, and increasing engagement	– Goal setting/ tracking – On-boarding – Team collaboration
2	Learning management systems	Although the company has a vast library of courses from Coursera and Harvard "MangeMentor", there is no clear evidence that these are automated or linked to an individual learning plan. Many valuable contents offered by KM team are not recorded and not searchable for future need	– Learning and development
3	Career development planning	Career development planning can be solved by planning a proper career path, identifying skills gaps, and monitoring capability enhancement and performance, in addition to offering an in-time compensation and career advancing opportunities for high performing individuals	– On-boarding – Career development planning – Learning and development – Succession planning – Compensation management – Competency management
4	Succession planning	Succession planning can be solved by implementing a TM software that comes with a succession planning component with a proper implementation plan	– Succession planning
5	Mobile-first	Having in mind a cloud-based system designed for mobile use can make a significant difference in its usability and effectiveness. Many TMSs come with a mobile version to ensure business continuity 24/7	– User experience – Mobile app
6	Data island	This issue can be solved by ensuring that the implemented system is integrated seamlessly with the other applications used by the information system of the targeted organization, such as microsoft dynamics	– System integration

Table 2 Evaluation of the available solutions based on 17 factors

Talent management features	BambooHR	UltiPro	Zoho Recruit	Ceridian Dayforce	Workable	iCIMS Recruit	CEIPAL TalentHire	Organimi	Cornerstone	Mentorcity	Monday.com	Talent Soft	Saba TMS	IBM Talent Management	Paycom	WorkforceNow	Oracle Taleo Cloud	Netchex	Microsoft Dynamics 365	Workday HCM	Qualtrics EmployeeXM	Total score
Starting price availability			1		1		1	1		1	1			1					1			1
Career development planning	2	2		2	2	2	2		2	2		2			2		2	2		2	2	2
Compensation management	2	2		2	2	2	2	2	2			2	2		2	2	2			2	2	2
Competency management	2	2		2	2	2	2	2	2	2					2	2	2	2		2	2	2
Employee lifecycle management	1	1		1	1	1	1		1			1		1	1	1		1		1	1	1
Goal setting / tracking	2	2		2	2	2	2		2	2		2	2		2		2	2	2	2	2	2
Job description management	1	1	1	1	1	1	1	1										1				1
On boarding	2	2	2	2	2	2	2	2	2	2	2	2	2		2	2	2	2	2	2	2	2
Performance management	1	1		1	1	1			1		1	1	1		1	1	1	1	1	1	1	1
Recruiting management	1	1	1	1	1	1	1	1	1		1	1	1	1	1	1	1	1	1		1	1
Succession planning		2		2	2	2		2	2	2		2	2				2			2	2	2
Learning and development	2	2		2	2	2		2	2	2	2	2	2		2		2	2	2	2	2	2
Employee wellbeing	1										1											1
Team collaboration				2								2							2			2
Gamification																			2			2
System integration	2	2	2	2	2	2	2	2									2		2			2
Mobile app	2	2	2	2	2	2	2	2	2		2	2	2			2		2	2	2		2
Score	75%	79%	32%	86%	82%	79%	64%	61%	68%	46%	36%	68%	50%	11%	54%	39%	64%	57%	61%	64%	61%	100%

The top five recommended TMS solution based on functionality matching the selection criteria is Ceridian Dayforce (86%), followed by Workable (83%), iCIMS Recruit (79%), UltiPro (79%), and BambooHR (76%).

5 Discussion

Since this research has not covered the actual testing and usage of the software, in reality, it might also be useful to consider another sort of recommendation. As different software review sites have different selection criteria, it might also be useful to consider them. For example, the reviewing site in [20] classifies high ranked TM apps based on five key factors: simplicity and ease of use, price and value for money, functionality, after-sales customer support, and likelihood to recommend. The reviews of [20] have also featured two additional software that were not included in the initial review: Bridge and Cezanne HR. One reason can be that Bridge does not categorize its products in the TMS category as their focus is more on learning management systems and performance management system solutions. Cezanne HR was not included in the initial study because the software does not use the term "talent management" in its website, hence it was not recognized by the search engine. Instead, Cezanne HR uses the term "people management" to describe TM functions. Another review in [14] offered an assessment on the top TM software based on popularity. Most solutions were included in our original ranking, except Oracle HCM Cloud. Instead, a different version more focused on TM (Oracle Taleo Cloud) was included. The reviews in [13] calculated

scores based on two pillars: Reviews and Web Search Interest. The following are the top five TM software based on Capterra's methodology [13]. TM Software in Capterra's TM are all already included in this study, except PrimePay which is more of a payroll software rather than a TM software, and it does not have most of the TM component features. The reviews in [12] listed 23 of the Leading TM Software in 2020. The website does not reveal their methodology or selection criteria, but they offer useful insight about each software including price, pros, and cons. There may be other significant factors to the targeted organization that this study did not explore such as security issues or on-premises cloud. Therefore, the organization might consider extending their Microsoft Dynamics solution services to cover HR and TM. It is essential to get a full briefing from the targeted organization to ensure that the selected solution is meeting the current and future need of the organization. Emerging TMS solutions such as BumbooHR and Ceridian Dayforce should also be examined, especially on what can be learnt from them.

6 Conclusion

The examined group of companies has achieved a major milestone in its HR practices among the top 2% companies in the world by receiving the "Invest in People" (IIP) Platinum Accreditation. Preserving this status requires continuous improvement of their HR practices with a clear focus on TM. This case study has examined the headquarter of a group of companies, evaluated and identified key areas related to TM for potential enhancement, provided alternative solutions and possible TMS mitigation tools, and concluded necessary steps needed to be taken in consideration before and during implementation of a TMS solution. This study is limited by the timeframe and limited access to information sources and qualitive methods. Future research can validate these findings by obtaining more detailed insights. These insights can be collected by conducting quantitative research and surveying the group's key stakeholders. Future studies can also benefit from the recent advancement in agile software development by focusing on implementing an agile TMS. Talents are highly connected to high performing organizations, where talents are considered the primary source of innovation and social development. TMS technologies not only help groups in the time of acquisition and merging to sustain their operations, they also create endless opportunities by unlocking the potential of their talented staff.

References

1. Mohammed AA, Gururajan R, Baig AH (2017) Primarily investigating into the relationship between talent management and knowledge management in business environment. Knowl Manage Bus Environ 2–8. https://doi.org/10.1145/3106426.31094441

2. Sivathanu B, Pillai R (2020) Technology and talent analytics for talent management—a game changer for organizational performance. Int J Organ Anal 28(2):457–473. https://doi.org/10.1108/IJOA-01-2019-1634
3. Collings DG, Mellahi K, Cascio WF (eds) (2017) The Oxford handbook of talent management. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198758273.001.0001
4. Salloum SA, Alshurideh M, Elnagar A, Shaalan K (2020) Machine learning and deep learning techniques for cybersecurity: a review. In: Joint European-US workshop on applications of invariance in computer vision. Springer, pp 50–57
5. Al-Maroof RS, Salloum SA, AlHamadand AQ, Shaalan K (2020) Understanding an extension technology acceptance model of google translation: a multi-cultural study in United Arab Emirates. Int J Interactive Mob Technol (IJIM) 14(03):157–178
6. AlShamsi M, Salloum SA, Alshurideh M, Abdallah S (2021) Artificial intelligence and blockchain for transparency in governance. Springer, pp 219–230
7. AlShehhi H, Alshurideh M, Kurdi BA, Salloum SA (2021) The impact of ethical leadership on employees performance: a systematic review. https://doi.org/10.1007/978-3-030-58669-0_38
8. Butler-Admas et al (2015) What it takes to be accredited INVESTORS IN PE PLE ® We invest in people framework
9. Vira R (2018) Gamification in the context of organizational talent management. In: 5th SGEM international multidisciplinary scientific conferences on SOCIAL SCIENCES and ARTS SGEM2018, Modern Science, STEF92 Technology, pp 439–446. https://doi.org/10.5593/sgemsocial2018/1.5/s05.055
10. Nwanisobi BC, Christopher IC (2020) Talent management: a conceptual framework with practical approach. Int J Recent Res Commerce Econ Manage (IJRRCEM) 7:1–10
11. Khdour N (2016) The reality of talent management implementation: a case study on Royal Jordanian Airlines. Int J Bus Manage 11(6):145. https://doi.org/10.5539/ijbm.v11n6p145
12. TechFunnel (2020) Top 23 talent management software in 2020 | TechFunnel, TechFunnel
13. Capterra (2019) Best talent management software 2021 | Reviews of the most popular tools & systems, Capterra
14. SOFTWARETESTINGHELP (2021) Top 10 talent management software systems in 2021 (reviews), Software testing help
15. Heathfield SM (2019) The best talent management practices. The Balance Careers
16. Lie Y, Sudyarti, Pardamean B (2017) Information system model of succession planning and career path. In: Proceedings of 2016 international conference on information management and technology, ICIMTech 2016, Institute of Electrical and Electronics Engineers Inc., pp 241–246. https://doi.org/10.1109/ICIMTech.2016.7930337
17. Mahjoub M, Atashsokhan S, Khalilzadeh M, Aghajanloo A, Zohrehvandi S (2018) Linking 'project success' and 'strategic talent management': satisfaction/motivation and organizational commitment as mediators. Proc Comput Sci, Elsevier B.V. 764–774. https://doi.org/10.1016/j.procs.2018.10.100
18. Murawski L (2020) Gamification in human resource management—Status quo and quo vadis. German J Human Resour Manage: Zeitschrift Für Personalforschung 239700222096179. https://doi.org/10.1177/2397002220961796
19. Stepper J (2016) The 5 elements of working out loud (revisited)—working out loud, Working Out Loud
20. GettApp (2020) Category leaders methodology | GetApp Research | GetApp®, GetApp

Adoption of Picture Archiving and Communication System (PACS) by Healthcare Professionals

Ibrahim Arpaci, Saeed Barzegari, and Fatemeh Askarian

Abstract The purpose of this study was to determine the factors affecting the acceptance of Picture Archiving and Communication System (PACS) by healthcare professionals based on the Unified Theory of Acceptance and Use of Technology (UTAUT). A total of 600 nurses, physicians, and paramedical staffs who perform their main tasks using the PACS in educational hospitals of Mazandaran University of Medical Sciences were participated in the study. Construct validity and internal reliability were tested by employing a confirmatory factor analysis (CFA) and a reliability analysis, respectively. Further, structural equation modeling (SEM) was employed to test hypothesized relationships. The results indicated that Performance Expectancy, Effort Expectancy, and Social Influence had a direct effect on Behavioral Intention. Whereas Facilitating Conditions and Behavioral Intention had a direct effect on Use Behavior.

Keywords PACS · Healthcare Professionals · UTAUT

1 Introduction

Technology has a major impact on social, economic, and political systems, and all of organizations, agencies, and individuals are increasingly required to use information technology in their own affairs [1]. The health system also benefits from the effects of information technology. One of the most widely used systems in healthcare has been the Picture Archiving and Communication System (PACS). PACS is a collection of technologies for acquisition, storage, and transmission of radiological images [2]. This system is one of the essential requirements for creating

I. Arpaci
Tokat Gaziosmanpasa University, 60250 Tokat, Turkey

S. Barzegari (✉)
Mazandaran University of Medical Sciences, Sari, Iran

F. Askarian
Shiraz University of Medical Sciences, Shiraz, Iran

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_63

an electronic health record. Advantages of PACS include accelerating image retrieval, improving image quality, remote reporting, effective communication between radiologists and physicians, facilitating the tasks and improving the accuracy of diagnosis and interpretation of images, linking images to the patients EHR, and reducing health care costs and medical errors. These advantages have led the PACS to have a crucial role in health care [3].

Of course, these benefits are achievable when implementing and using this technology is precisely and accurately. One of the most important barriers to PACS implementation, besides financial issues, is human barriers [4]. However, this technology has been slowly accepted by healthcare providers, and therefore, identifying and measuring the effective factors on its acceptance, and attempting to strengthen the positive factors and providing solutions to overcome the deterrent factors will have a major role in facing barriers to its acceptance [5].

Venkatesh et al. [6] presented the Unified Theory of Acceptance and Use of Technology (UTAUT) for identifying the factors influencing technology acceptance. Four factors are included in this model: Performance Expectancy, Effort Expectancy, Social Influence and Facilitating Conditions and gender, age, and experience considered as moderating factors [7]. According to Venkatesh et al. [8], Performance Expectancy is a degree to which the users expects that using the system will help them to attain gains in job performance. This factor consists of five structures: perceived usefulness, outcome expectations, job-fit, relative advantage, and extrinsic motivation. Effort Expectancy is the degree of ease associated with the use of the system [9]. Social Influence is the degree to which a user perceives that important others believe he or she should use the new system [10]. Facilitating Conditions are the degree to which a user believes that an organization and technical infrastructure exists to support use of the technology [11].

This model, which is a combination of individual and organizational constructs, has been tested in different countries and has provided acceptable results. Considering the significant role of the PACS in improving the quality of care services, the purpose of this study was to determine the effective factors in the acceptance of PACS by physicians, nurses, and paramedics in educational hospitals of Tehran University of Medical Sciences (TUMS) based on the UTAUT.

2 Method

In this cross-sectional study, study population includes 600 nurses, physicians, and paramedical staffs of 3 educational hospitals of TUMS that perform their entire or part of tasks with PACS system. The PACS system in these hospitals was responsible for recording, collecting, integrating, reporting, and transmitting patient's therapeutic and diagnostic data. Employees who use PACS exclusively for financial affairs did not participate in this study. A total of 600 nurses, physicians, and paramedical staff randomly participated in the study. We used a questionnaire consisting of two parts to collect data. The first part was a referral and the purpose

of the data collection and the necessity of the respondent's cooperation to complete the questionnaire was stated. In the second section, there were demographic questions and the UTAUT questionnaire items. UTAUT is a five-point Likert scale ranged from one "Completely Disagree" to five "Completely Agree". The validity and reliability of the Persian version of UTAUT have been confirmed in Najafi et al. study [12].

Goodness of the fitting of the factors, examined by confirmatory factor analysis (CFA). Various fitting indicators such as $\chi^2/df < 3$, RMSEA < 0.06, CFI > 0.90, GFI > 0.90, IFI > 0.90, TLI > 0.90, NFI > 0.90, and RMR < 0.5 were studied. Cronbach's alpha coefficient equal to and higher than 0.7 was used to accept internal consistency [13]. Then, 5 hypotheses of the study were tested by structural equation modeling (SEM) and the goodness of fitting was examined with indexes that mentioned for CFA.

Hypothesis 1. The Performance Expectancy has a direct effect on the Behavioral Intention.
Hypothesis 2. The Effort Expectancy has a direct effect on the Behavioral Intention.
Hypothesis 3. The Social Influence has a direct effect on the Behavioral Intention.
Hypothesis 4. The Facilitating Conditions have a direct effect on the Use Behavior.
Hypothesis 5. The Behavioral Intention has a direct effect on the Use Behavior.

3 Results

The characteristics of the participants in the study were (from 673 requests) 600 (318 nurse, 221 paramedical staff, 61 doctor) users with an average age of 35.69 ± 5.41, most of them were 61% married, 84% of the undergraduate degree, an average working experience of 13 years, 33% of the participants from the Shahid Rajaee Cardiology Hospital, 34% of the participants Khatam Al-Anbia Private Hospital, 33% of Valiasr Hospital Participants.

Confirmatory factor analysis has been used for assessing construct validity of tool. The factor loading of items 8 (0.23), 10 (0.33), 17 (0.39), and 18 (0.34) were less than 0.40, therefore, these items were deleted. The explained variance by these items was 66.91. The value obtained for the goodness of fit indicators presented acceptable χ^2/df (3.37), df (168), p-value (≤ 0.001), RMSEA (0.053), CFI (0.944), GFI (0.940), IFI (0.945), TLI (0.930), NFI (0.923), and RMR (0.030).

The internal consistency of the questionnaire was evaluated by Cronbach's alpha test and for the total questionnaire it is 0.871, and for performance expectancy (0.790), effort expectancy (0.862), social influence (0.747), facilitation conditions (0.815), behavioral intention (0.830), use behavior (0.706), all of which are higher than 0.70 and represent a high internal consistency of the questionnaire. The factor loading and internal consistency of items presented in Table 1.

Table 1 Factor loading and internal consistency of items

Construct	Item	Mean	Skewness	Kurtosis	Factor loading	α
Performance expectancy	Q1	3.99 ± 0.9	−0.877	0.902	0.73	0.76
	Q2	3.27 ± 1.1	0.037	−0.745	0.88	0.75
	Q3	3.88 ± 1	−0.689	0.017	0.56	0.78
	Q4	4.09 ± 1	−0.998	0.279	0.59	0.78
Effort expectancy	Q5	3.81 ± 1.2	−0.666	−0.486	0.82	0.85
	Q6	3.32 ± 1	0.035	−0.704	0.83	0.79
	Q7	3.79 ± 1	−0.429	−0.600	0.81	0.83
Social influence	Q9	3.28 ± 1.3	−0.249	−0.920	0.54	0.73
	Q11	2.92 ± 1.1	0.530	−0.671	0.71	0.74
	Q12	3.07 ± 1.2	0.319	−1.111	0.83	0.72
Facilitating conditions	Q13	3.80 ± 1.1	−0.733	−0.266	0.66	0.80
	Q14	3.77 ± 1.1	−0.743	−0.296	0.69	0.78
	Q15	3.32 ± 1.1	−0.047	−0.644	0.79	0.77
	Q16	3.74 ± 1.2	−0.707	−0.589	0.57	0.79
Use behavior	Q19	3.80 ± 1.1	−0.435	-1.058	0.60	0.66
	Q20	4.25 ± 0.9	−1.310	1.435	0.69	0.67
	Q21	4.08 ± 1	−0.964	0.553	0.68	0.69
	Q22	3.76 ± 1.1	−0.578	−0.619	0.87	0.70
Behavioral intention	Q23	3.96 ± 1.1	−0.764	−0.749	0.85	0.78
	Q24	3.73 ± 1.1	−0.319	−0.985	0.45	0.79
	Q25	3.98 ± 1.1	−1.133	0.852	0.56	0.80

Finally, using structural equation modeling, we tested the hypotheses of the study and direct and indirect effects of the factors on Use Behavior. The conceptual model is also shown in Fig. 1. The fitting results of the model was presented acceptable χ^2/df (3.49), df (255), p-value (≤ 0.001), RMSEA (0.054), CFI (0.944), GFI (0.937), IFI (0.942), TLI (0.927), NFI (0.920), and RMR (0.031). The total variance explained was equal to 62%. The results revealed the direct effect of Performance Expectancy (t-value = 1.392, Direct effect = 6.5, p-value = 0.164), Effort Expectancy (t-value = 2.637, Direct effect = 12.1, p-value = 0.008), and Social Influence (t-value = 4.317, Direct effect = 20.3, p-value = ≤ 0.001) factors on Behavioral Intention, and also direct effect of Facilitating Conditions (t-value = 2.847, Direct effect = 10.4, p-value = 0.004) and Behavioral Intention (t-value = 11.217, Direct effect = 77.1, p-value = ≤ 0.001) on the Use Behavior. Also, the results indicated that performance expectancy (indirect effect = 5, p-value = ≤ 0.001), effort expectancy (indirect effect = 9.2, p-value = ≤ 0.001), and social influence (indirect effect = 15.5, p-value = ≤ 0.001) have significant indirect effects on Use Behavior.

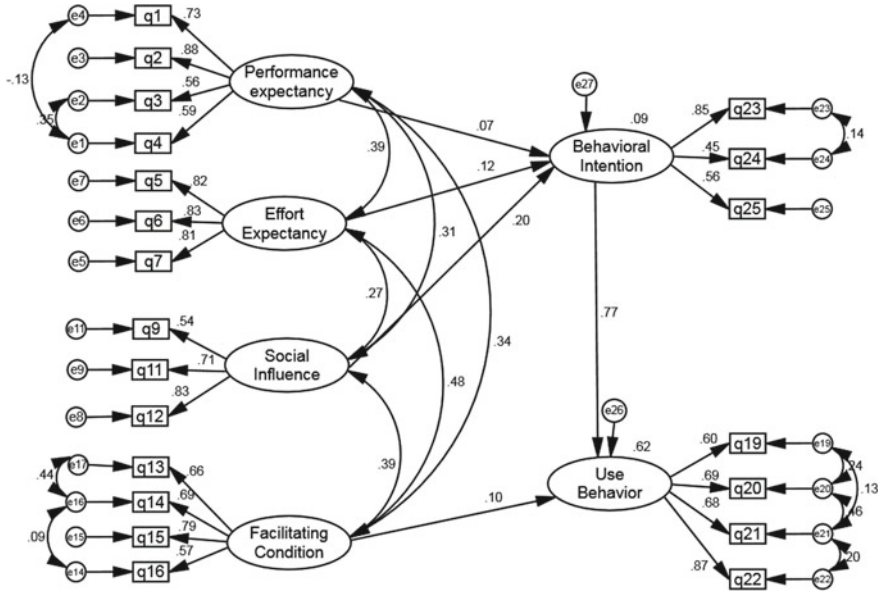


Fig. 1 Structural model of the study

4 Discussion and Conclusion

Picture Archiving and Communication System (PACS) has a crucial role in health care. Identifying and measuring the effective factors on its acceptance, and attempting to strengthen the positive factors and providing solutions to overcome the deterrent factors will have a major role in facing barriers to its acceptance. The purpose of this study was to determine the effective factors in the acceptance of PACS by physicians, nurses, and paramedics in educational hospitals of Mazandaran University of Medical Sciences based on the Unified Theory of Acceptance and Use of Technology (UTAUT).

In this cross-sectional study, 600 nurses, physicians, and paramedical staffs that perform their entire or part of tasks with PACS system completed the UTAUT Questionnaire. Construct validity and internal reliability examined using confirmatory factor analysis (CFA) and Cronbach's alpha coefficient, respectively. Then, direct effect of Performance Expectancy, Effort Expectancy, and Social Influence factors on Behavioral Intention, and also direct effect of Facilitating Conditions and Behavioral Intention on the Use Behavior were tested by structural equation modeling (SEM). The results indicated that Performance Expectancy, Effort Expectancy, and Social Influence had a direct effect on Behavioral Intention. Further, Facilitating Conditions and Behavioral Intention had a direct effect on Use Behavior.

Previous findings confirmed our findings. For example, Duyck et al. [14] investigated adoption of a PACS by applying the UTAUT in radiological settings. Their findings indicated that Performance Expectancy and Facilitating Conditions were significant for predicting Behavioral Intention to use PACS. In another study, Garavand et al. [15] investigated user acceptance of PACS based on Technology Acceptance Model. Their findings indicated that perceived usefulness and ease of use were significantly associated with use of PACS.

The present study applied a well-structured model (i.e., UTAUT) to predict user acceptance of PACS by physicians, nurses, and paramedics in educational hospitals of Mazandaran University of Medical Sciences. One of the limitations of the study was cross-sectional nature of the research. Future studies should therefore extend the research model by additional constructs and collect longitudinal data to better understand user behavior.

Funding

This work was supported by the Mazandaran University of Medical Sciences with code IR.MAZUMS.REC.1397.1202.

References

1. Locatelli G, Mancini M, Romano E (2014) Systems engineering to improve the governance in complex project environments. Int J Project Manage 32(8):1395–1410
2. Paré G, Lepanto L, Aubry D, Sicotte C (2005) Toward a multidimensional assessment of picture archiving and communication system success. Int J Technol Assess Health Care 21(4):471–479
3. Mathieson K (1991) Predicting user intentions: comparing the technology acceptance model with the theory of planned behavior. Inf Syst Res 2:173–191
4. Piette JD, Lun K, Moura LA Jr, Fraser HS, Mechael PN, Powell J et al (2012) Impacts of e-health on the outcomes of care in low-and middle-income countries: where do we go from here? Bull World Health Organ 90(5):365–372
5. Sittig DF, Singh H (2010) A new sociotechnical model for studying health information technology in complex adaptive healthcare systems. Qual Saf Health Care 19(3):i68–i74
6. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425–478
7. Kijsanayotin B, Pannarunothai S, Speedie SM (2009) Factors influencing health information technology adoption in Thailand's community health centers: applying the UTAUT model. Int J Med Informatics 78(6):404–416
8. Shibl R, Lawley M, Debuse J (2013) Factors influencing decision support system acceptance. Decis Support Syst 54(2):953–961
9. Duyck P, Pynoo B, Devolder P, Voet T, Adang L, Vercruysse J (2008) User acceptance of a picture archiving and communication system-applying the unified theory of acceptance and use of technology in a radiological setting. Methods Inf Med 47(2):149–156
10. Zuiderwijk A, Janssen M, Dwivedi YK (2015) Acceptance and use predictors of open data technologies: drawing upon the unified theory of acceptance and use of technology. Gov Inf Q 32(4):429–440
11. Arpaci I (2015) A comparative study of the effects of cultural differences on the adoption of mobile learning. Br J Edu Technol 46(4):699–712

12. Najafi N (2010) Effective factors for adoption of electronic patient record between physician using the unified theory of acceptance and use of technology (UTAUT). Payam Nour, Shiraz
13. Cronbach LJ (1951) Coefficient alpha and the internal structure of tests. Psychometrika 16(3):297–334
14. Duyck P, Pynoo B, Devolder P, Voet T, Adang L, Vercruysse J (2008) User acceptance of a picture archiving and communication system. Methods Inf Med 47(2):149–156
15. Garavand A, Ghanbari S, Ebrahimi S, Kafashi M, Ahmadzadeh F (2015) The effective factors in adopting picture archiving and communication system in Shiraz educational hospitals based on technology acceptance Model. J Health Biomed Inf 1:76–82

Robotic Process Automation Technology Acceptance: A Case Study from Austrian Companies

Martin Tschandl, Vitaliy Mezhuyev, and David Möstl

Abstract Robotic Process Automation (RPA) has become an important technology in the context of digitalization and automation of business processes. RPA is more flexible than stable business process automation of the IT systems at the user interface (UI) level, which leads to a sustainable reduction in process costs. However, the acceptance of RPA technology in Austrian companies is still very low. The problem may refer to the shortage of perceived maturity of RPA technology and/or the inertia of the industry. Thus, this paper describes an empirical study that evaluates the attitudes of Austrian companies concerning the use of RPA in industrial processes. The theoretical basis for the evaluation is the Technology Acceptance Model (TAM), which is expanded by the factor of perceived maturity. After the collection of data from 251 companies, the Partial Least Squares Structural Equation Modelling (PLS-SEM) was conducted to assess the measurement and structural models. The study showed that perceived usefulness and perceived maturity have an important influence on RPA technology acceptance. At the same time, there is no such effect for the perceived ease of use. The R2 was 0.481 for RPA acceptance, which points out that the explanatory power of the proposed model is relatively substantial. The empirical results have implications for the managers responsible for employing RPA technology in their companies.

Keywords Robotic process automation · Technology acceptance model · PLS-SEM · Austrian companies

M. Tschandl · V. Mezhuyev (✉) · D. Möstl
FH Joanneum University of Applied Sciences, Werk-VI-Straße 46, 8605, Kapfenberg, Austria
e-mail: vitaliy.mezhuyev@fh-joanneum.at

M. Tschandl
e-mail: martin.tschandl@fh-joanneum.at

D. Möstl
e-mail: david.moestl@edu.fh-joanneum.at

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_64

1 Introduction

Automation in the administrative domain began in the 1990s with *Straight Through Processing* (STP) [1]—e.g., the automated copying of data—moved on to the large area of *Business Process Automation* (BPA), specifically reengineering and automation of processes with operational software solutions and stable interfaces—e.g., workflow management or Enterprise Resource Planning (ERP) [2], and finally came *to Robotic Desktop Automation* (RDA) for automation in (only) one application and *Robotic Process Automation* (RPA) in multiple applications. Some authors distinguish RPA between Click-agents (which use the same interfaces as human users) and Dialogue-agents (which can be addressed in the same way as human colleagues, e.g., text/voice-based chatbots) [3].

The limitations of RDA—it does not detect errors itself, created automations cannot be reused for other processes in the company, and startup requires manual triggering [4]—do not apply to RPA. Rather, the software robot starts autonomously as soon as predefined events occur. The underlying automated IT systems and applications do not recognize that they are being served by a software robot, which only accesses the user interface or the presentation level, whereas STP and BPA directly address the data management level [5]. Given in Table 1 a systematic analysis of RPA definitions by 13 authors from 2016 to 2020 leads to the following definitional characteristics used by the majority of these authors [4–16]: RPA is a software program that mimics repetitive tasks previously performed by humans on computer systems and is used across applications. It is therefore also referred to as a virtual employee or software robot [12]. RPA executes decisions in a structured and rule-based manner based on user interface interventions; no software interfaces need to be programmed. Thus, RPA is trained rather than programmed, resulting in low implementation time. Artificial intelligence is not integrated at this stage of development.

The next stage of development, *Intelligent Process Automation* (IPA) includes artificial intelligence and will grow strongly, especially in the financial sector [17].

In business practice, RPA technology is typically applied to use cases such as the following [3]: HR onboarding, assignment of credit card items to trips in travel expense reimbursement processes, review of open orders, mail-to-ticket-processes in service centers, data collection from various systems and compilation of reports, master data updating/integration, and duplicate checks.

The broad applicability is associated with three general potential benefits: Cost savings, process quality improvements, and time savings [5, 15, 16]. The *cost savings* results from the difference of the personnel hours and thus costs for the previous business processes with those of the implementation and operation of the bots. Since only minor changes in the IT infrastructure are required for an RPA implementation, the development of an initial virtual bot is possible within two to four weeks [18]. These savings result in a high return on investment [11] and a relatively low amortization period of approximately twelve months, including all relevant costs of the virtual robot [9]. The *process quality improvements* result from

Table 1 Definition elements of robotic process automation in literature

	Kirchberg et al. 2016, p. 87ff	Allweyer 2016	Willcocks et al. 2017, p. 2ff	Manutiu 2018, p. 5ff	Alexander et al. 2018, p. 12f	Peper 2018, p. 20ff	Smeets, Erhard, Kaußler 2018, p. 7ff	Van der Aalst et al. 2018, p. 269ff	Hermann, Stoi, Wolf 2018, p. 28ff	Hofmann, Samp, Urbach 2019, p. 2ff	Doguc 2020, p. 469ff	Langmann, Duri 2020, p. 5ff	Radke, Dang, Tan 2020, p. 130	Number of mentions
Execute decisions in a structured and rule-based manner	x	x	x	x	x	x	x	x	x	x	x	x	x	13
No interfaces necessary (cross-application use)	x	x	x	x	x	x	x	x	x	x	x	x	x	13
Imitation of human activity	x	x		x	x	x	x	x	x	x		x	x	11
User interface intervention (inside-out approach)		x		x	x		x	x	x	x	x	x	x	10
Low implementation time			x	x		x	x	x	x	x	x	x	x	10
Use in repetitive tasks			x		x	x		x	x	x	x	x	x	9
No integration of artificial intelligence		x		x	x		x			x		x	x	7
Trained instead of programmed (no programming knowledge necessary)		x	x				x			x	x	x	x	7

the reduction in the failure rate [1] for repetitive human tasks [5]. The *time savings* in process runtime due to the speed of virtual robots potentially leads to an improvement in customer satisfaction, both for internal and external process customers [8].

Despite these potential benefits for many processes, the penetration of companies with RPA technology is not yet far advanced. This can be seen, for example, in a study of Austrian companies, in which only 15% stated that they already use RPA technology, of which 32% are large companies (>250 employees) and only 7% are Small and Medium-Sized Enterprises (SMEs, $\leq$250 employees). At the same time, companies have a limited view of the importance (important/very important) of RPA, with only 31% of large companies and 15% of SMEs stating this, while they will state the opportunities (chance/very chance) relatively high with 83% and 55%, respectively [19]. Obviously, there is (still) an acceptance problem with RPA, which delays a faster penetration of companies with this technology. This paper aims to use an adapted TAM approach to explain the acceptance gap with the data of a study among Austrian companies.

2 Literature Review

Information technology, whereas RPA is a new example, has the potential to gain a sustainable cost reduction efficiency but the user's unwillingness to accept the technology hinders its potential [20]. Therefore, Davis wanted to test the acceptance of new information technologies by its users [21].

The starting point for the Technology Acceptance Model (TAM) by Davis was the Theory of Reasoned Action (TRA) [22] and the Theory of Planned Behaviour (TPB) [23]. TRA has proven success in predicting and additionally in the explanation of human behaviour across different fields [24]. TPB expands the TRA by the factor of social actions that follow a kind of developed plan [23]. Davis combined both theories, and the proposed model achieved a prediction power to answer why users accept or do not accept a particular information system [25]. Since the development, TAM was expanded and applied to other technologies, which lie outside of the information system domain [21]. TAM has already been successfully applied in different areas like search-based software engineering [26], mobile payment [27], metamodeling [28], and RPA [29] itself. TAM was also used to evaluate a software process improvement initiative and, thus, can measure accepting a new technology as the reasons are similar for an initiative and a new technology [30].

Following TAM, the two main factors, Perceived Ease of Use (PEOU) and Perceived Usefulness (PU), affect the acceptance of new technologies. PEOU is described as "the degree to which a person believes that using a particular system would be free from effort" [24]. The effort from a person is finite and can only be given to a certain number of activities. Therefore, the user's acceptance is more likely if a user feels more comfortable with a system vs other options [24]. PU

designates "the degree to which a person believes that using a particular system would enhance his or her job performance" [24]. If a system is useful, the system can be used more advantageously, and hence a user believes in a higher performance increase for him [24].

To summarize, if a user believes that the system usage is easy, with a low amount of effort, and the system increases the accomplishment of application results, the user is more likely to accept the technology. This premise is the general assumption of TAM.

TAM was already applied to check an acceptance of the RPA technology, where Wewerka et al. [29] used a classical TAM model. In contrast to the model proposed by [29], this paper expands the classical TAM by the factor of Perceived Maturity (PM). PM is "the degree to which tools are perceived as mature and suitable for the tasks in hand" [31]. The reason for the use of the PM factor is that RPA is a relatively new software technology, where maturity is significant to explain and predict a technology acceptance.

3 Proposed Model

The literature study was conducted to identify the factors which may influence the acceptance of RPA technology. The proposed model includes perceived usefulness (PU) and perceived ease of use (PEOU) of classical TAM, which is expanded by the perceived maturity (PM) as an additional factor. Figure 1 illustrates the proposed research model. It supposes that acceptance of RPA technology is influenced by three factors, specifically, PU, PEOU, and PM, where PU is also affected by PEOU.

Therefore, we formulate the following hypotheses:

H1: Perceived usefulness will positively affect RPA technology acceptance.
H2: Perceived ease of use will positively affect RPA technology acceptance.
H3: Perceived maturity will positively affect RPA technology acceptance.
H4: Perceived ease of use will positively affect perceived usefulness.

4 Research Methodology

4.1 Instrument Development

The participants of this study are employees working in different companies in various industries. All considered companies are located in Austria and have more

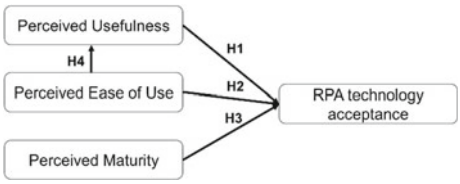


Fig. 1 Proposed model

than ten employees. An online questionnaire survey was developed with the Lime survey tool and next sent through email for data gathering. The survey comprised of three parts, where the first part collects the employees' data like their position in the company. Additionally, data from the companies like the revenue, the number of employees, and the legal form were collected. The second part measures the respondents' attitudes towards using RPA technology based on the proposed model. The questionnaire also includes the open answer section for evaluation of chances, risks, and the types of RPA usages in general.

4.2 Methods of Data Analysis

This research followed the guidelines for performing information systems adoption research [32]. The PLS-SEM method employing SmartPLS tool was applied for making the model evaluation [33]. To validate the measurement model, researchers analyse the items outer loadings and also the average variance extracted (AVE) [34]. For confirming a discriminant validity, the cross-loadings, the Fornell-Larcker criterion [34] and the Heterotrait-Monotrait ratio (HTMT) were calculated [35]. The path coefficients and the R2 coefficient [34] were measured for the structural model assessment.

5 Results and Discussion

5.1 Descriptive Analysis

This part of the study describes the demographic data of the employees and their companies. Of the 251 participants, 30% worked in industry, 20% in craft, 12% in trade, 9% in tourism, 7% information/consulting, 6% in transport, 5% in banks, and 10% in other sectors. The companies were classified by the number of employees, whereas 66% were SME, and 34% were large companies. 76% of the respondents are in management positions consisting the executive management 44%, the head of departments 26%, and authorized representatives 6%. The other 24% consists of 14% employees, 8% assistant to executive management, and 2% are in other positions.

5.2 Measurement Model Assessment

To measure the reliability of each item, their loadings were calculated. Following [34], a value ≥ 0.7 for the loading of each item is counted as reliable. The same rule applies to the values of Cronbach's alpha and composite reliability. Table 2 shows that all the items meet the suggested criteria and thus are reliable.

Table 2 Construct reliability and validity

Constructs	Cronbach's alpha	Items loadings	Composite reliability	AVE
PEOU	0.722	0.938 0.817	0.872	0.773
PM	0.799	0.804 0.841 0.880	0.880	0.709
PU	0.903	0.844 0.897 0.911 0.865	0.932	0.774
RA	0.749	0.862 0.923	0.887	0.797

The AVE is the main measure for confirming the convergent validity, which is calculated as the grand mean value of the squared loadings of the items related to the construct. An AVE value $\geq$0.5 denotes that the construct explains >0.5 of the variance of its items [34]. Table 2 shows that the values of AVE are >0.5 and all the values of composite reliability and Cronbach's alpha are >0.7. Hence, convergent validity is confirmed.

The Fornell-Larcker criterion, the HTMT ratio, and cross-loadings were measured to endorse the discriminant validity. The square root of the AVE values (marked by bold in Table 3) exceed the correlation of the constructs, which confirms the Fornell-Larcker criterion.

Table 4 shows that the loading of every item is more than the loadings of corresponding constructs' items, thus confirming the cross-loadings criterion.

Table 5 confirms the discriminant validity of the model, showing that all computed HTMT ratio values are less than 0.85.

5.3 Assessment of the Structural Model

In the given study, R2 and the path measurements were taken as the main criteria for the assessment of the structural model [34]. Figure 2 shows that the value of R2 is 48.1% for RPA technology acceptance and 4.7% for PU.

In social science research, R2 values of 0.75, 0.50, or 0.25 can be respectively described as substantial, moderate, and weak [6]. Since the R2 indicates the amount of a shared variation between two or more variables, we can say about a moderate prediction level of RPA technology acceptance by PU, PEOU, and PM factors.

Figure 2 together with Table 6 describe the path measurements M and p-values for the formulated hypothesis. The paths between variables are significant in three of four cases, which reveals that three of four formulated hypotheses are supported.

Table 3 Fornell-larcker criterion

	PEOU	PM	PU	RA
PEOU	**0.879**			
PM	0.469	**0.842**		
PU	0.218	0.494	**0.880**	
RA	0.376	0.587	0.593	**0.893**

Table 4 Cross loadings

	PEOU	PM	PU	RA
PEOU1	**0.938**	0.500	0.264	0.373
PEOU2	**0.817**	0.284	0.078	0.272
PM1	0.524	**0.804**	0.282	0.401
PM2	0.374	**0.841**	0.336	0.446
PM3	0.331	**0.880**	0.571	0.600
PU2	0.175	0.400	**0.897**	0.466
PU3	0.165	0.433	**0.911**	0.496
PU4	0.135	0.367	**0.865**	0.542
RA1	0.424	0.510	0.353	**0.862**
RA2	0.270	0.538	0.668	**0.923**
PU1	0.275	0.519	**0.844**	0.565

Table 5 Heterotrait-monotrait ratio

	PEOU	PM	PU	RA
PEOU				
PM	**0.605**			
PU	**0.234**	**0.546**		
RA	**0.512**	**0.738**	**0.687**	

H1 ($M = 0.404$, $p < 0.05$) explains the path connecting PU and RPA technology acceptance, showing that PU positively influences the acceptance of RPA technology by Austrian companies. H2 ($M = 0.136$, $p > 0.05$) reveals the path connecting the PEOU and RPA technology acceptance, indicating that the PEOU factor is not significant for the acceptance of RPA technology. H3 ($M = 0.324$, $p < 0.05$) depicts the path connecting the PM and RPA technology acceptance, representing that the PM boosts the acceptance of the RPA technology. Finally, H4 ($M = 0.218$, $p < 0.05$) shows the path connecting the PEOU and the PU, disclosing that the PEOU positively influences the PU of RPA technology.

The result of this study suggests that the PU and PM positively influence the RPA technology acceptance by Austrian companies. At the same time, there is no such effect discovered for the PEOU factor.

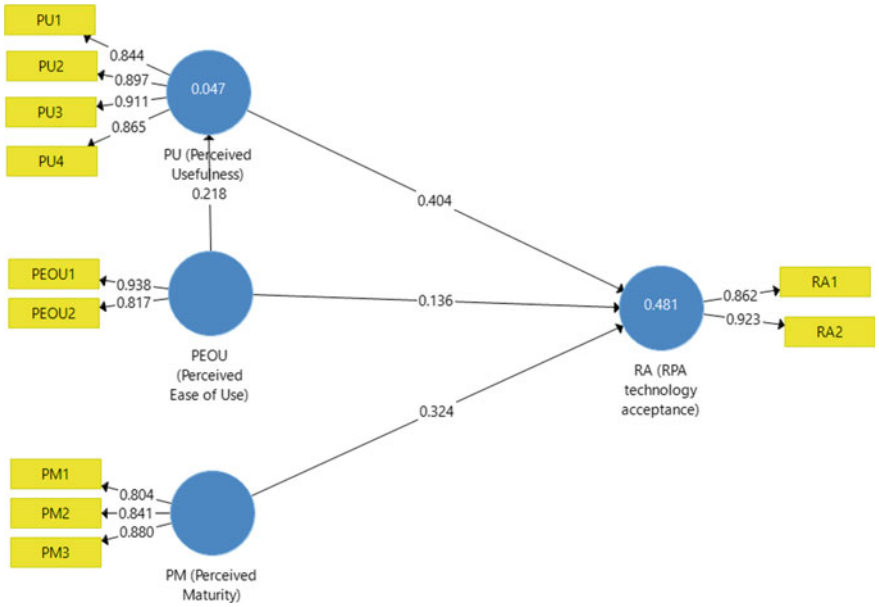


Fig. 2 Evaluation of the proposed RPA-TAM in SmartPLS

Table 6 Hypotheses test results

Paths	Hypotheses	Path coefficients	P values	Results
PU -> RA	H1	0.404	0.000	Supported
PEOU -> RA	H2	0.136	0.071	Not supported
PM -> RA	H3	0.324	0.000	Supported
PEOU -> PU	H4	0.218	0.020	Supported

6 Conclusion

This study gives two contributions. First, the paper examines the RPA technology acceptance in Austrian companies focusing on the general industry with the exemption of public institutions and with a minimum of ten employees. Second, to examine the RPA technology acceptance, the TAM was extended by the perceived maturity factor. Our results indicated that perceived usefulness and perceived maturity significantly influence an acceptance of RPA technology. Furthermore, perceived ease of use was discovered to influence the perceived usefulness of RPA technology. At the same time, the hypothesis about the significant effect of perceived ease of use on RPA technology acceptance was not supported. In the contrast to classical TAM model, employees of Austrian companies do not consider the easiness of use as a decisive factor for the RPA technology acceptance. This can be

considered as a contribution to the existing body of scientific knowledge. The empirical results also have administrative implications: the executives should consider these factors while implementing RPA technology in their companies.

The limitation of the proposed model is that TAM was extended by only one factor, namely, the perceived maturity. Future work will examine additional factors which may influence RPA technology acceptance. The comparison of companies of different types and number of employees will be also done. Furthermore, other than Austria countries will be considered in the future empirical study to deduce a possible difference in responses.

Appendix A. The Survey Questions

Perceived Usefulness

PU1: Robotic Process Automation can improve the process quality.
PU2: Robotic Process Automation can accelerate the process execution.
PU3: Robotic Process Automation can save time.
PU4: Robotic Process Automation can improve productivity.

Perceived Ease of Use

PEOU1: Robotic Process Automation is easy to use.
PEOU2: Robotic Process Automation requires no training.
PEOU3: Robotic Process Automation is practical.

Perceived Maturity

PM1: Robotic Process Automation is mature and reliable.
PM2: Experiences already exist for Robotic Process Automation.
PM3: Robotic Process Automation has many application possibilities.

RPA technology acceptance

RA1: I like to use RPA.
RA2: I have a positive attitude towards RPA technology.

References

1. Osman C-C (2019) Robotic process automation: lessons learned from case studies. Inform Econ 23:66–75
2. Watson E, Holmes K (2009) Business process automation. In: Nof S (ed) Handbook of automation. Springer, Berlin Heidelberg, pp 1597–1612
3. Schulze M, Nuhn H (2020) Process automation in controlling—results of an empirical study. ICG, St. Gallen
4. Willcocks L, Lacity M (2017) Robotic process automation: strategic transformation lever for global business services? J Inf Technol Teach Cases 7:17–28
5. Smeets M, Erhard R, Kaußler T (2019) Robotic process automation (RPA) in der Finanzwirtschaft. Technologie—implementierung—erfolgsfaktoren für entscheider und anwender. Springer Gabler, Wiesbaden
6. Kirchberg A, Müller D (2016) Digitalisierung im controlling: einflussfaktoren, standortbestimmung und konsequenzen für die controllerarbeit. In: Gleich R, Grönke K, Kirchmann M, Leyk J (eds) Konzerncontrolling, pp 79–96. Haufe, Freiburg
7. Allweyer T (2020) Robotic process automation—neue perspektiven für die prozessautomatisierung. http://www.kurze-prozesse.de/blog/wp-content/uplo-ads/2016/11/Neue-Perspektiven-durch-Robotic-Process-Automation.pdf. Last Accessed on 2020/02/07
8. Manutiu S (2018) Digitalisierung im controlling—mehrwert durch robotic process automation. Controlling 30:4–10
9. Alexander S, Haisermann A, Schabicki T, Frank S (2018) Robotic process automation (RPA) im rechnungswesen und controlling—welchen chancen ergeben sich. Controlling 30:11–20
10. Peper I (2018) Effizienzsteigerungen im controlling durch robotic process automation. Controlling 30:20–27
11. Van der Aalst W, Bichler M, Heinzel A (2018) Business process management: a survey. Bus Inf Syst Eng 60:269–272
12. Herman K, Stoi R, Wolf B (2018) Robotic process automation im finance and controlling der MANN+HUMMEL Gruppe. Controlling 30:28–34
13. Hofmann P, Samp U, Urbach N (2018) Robotic process automation. Electron Mark 30:1–8
14. Doguc O (2020) Robot process automation (RPA) and its future. In: Hacioglu U (ed) Handbook of research on strategic fit and design in business ecosystems, pp 469–492. IGI Global, Hershey
15. Langmann C, Turi D (2020) robotic process automation (RPA)—digitalisierung und automatisierung von prozessen. Springer Gabler, Wiesbaden
16. Radek AM, Dang MT, Tan A (2020) Using robotic process automation (RPA) to enhance item master data maintenance process. Sci J Logist 16:129–140
17. Cole T (2020) Erfolgsfaktor Künstliche Intelligenz. Carl Hanser, München
18. Asatiani A, Penttinen E (2016) Turning robotic process automation into commercial success—case OpusCapita. J Inf 6:67–74
19. Tschandl M, Möstl D, Mezhuyev V (2020) Robotic process automation in der praxis: nutzen und grenzen von software-robotern. WINGbusiness. 40:35–39
20. Bowen, W.: The Puny Payoff from Office Computers. Fortune Mag. 20–24 (1986).
21. Davis FD, Bagozzi RP, Warshaw RR (1989) User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 35:989–1003
22. Ajzen I, Fishbein M (1980) Understanding attitudes and predicting social behavior. Prentice-Hall
23. Ajzen I (1985) From intentions to actions: a theory of planned behavior. In: Kuhl J, Beckmann J (eds) Action control: from cognition to behavior. Springer, Berlin Heidelberg, pp 11–39
24. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13:319–340

25. Turner M, Kitchenham B, Brereton P, Charters S, Budgen D (2010) Does the technology acceptance model predict actual use? A systematic literature review. Inf Softw Technol 52:463–479
26. Mezhuyev V, Al-Emran M, Ismail MA, Benedicenti L, Chandran D (2019) The acceptance of search-based software engineering techniques: an empirical evaluation using the technology acceptance model. IEEE Access 7:101073–101085
27. Li J, Wang J, Wangh S, Zhou Y (2019) Mobile payment with Alipay: an application of extended technology acceptance model. IEEE Access 7:50380–50387
28. Mezhuyev V, Al-Emran M, Fatehah M, Hong NC (2018) Factors affecting the metamodeling acceptance: a case study from software development companies in Malaysia. IEEE Access 6:49476–49485
29. Wewerka J, Dax S, Reichert M (2020) A user acceptance model for robotic process automation. 2020 IEEE 24th Int Enterp Distrib Object Comput Conf EDOC 97–106
30. Umarji M, Seaman C (2005) Predicting acceptance of software process improvement. ACM Sigsoft Softw Eng Notes 30:1–6
31. Mohagheghi P, Gilani W, Stefanescu A, Fernandez MA (2013) An empirical study of the state of the practice and acceptance of model-driven engineering in four industrial cases. Empirical Softw Eng 18:89–116
32. Al-Emran M, Mezhuyev V, Kamaludin A (2018) PLS-SEM in information systems research: a comprehensive methodological reference. In: 4th International conference on advanced intelligent systems and informatics (AISI 2018). Springer (In Press)
33. Ringle CM, Wende S, Becker J-M (2015) SmartPLS 3. Bönningstedt: SmartPLS
34. Hair JF, Hult GTM, Ringle C, Sarstedt M (2014) A primer on partial least squares structural equation modeling (PLS-SEM). Sage Publications
35. Henseler J, Ringle CM, Sarstedt M (2015) A new criterion for assessing discriminant validity in variance-based structural equation modeling. J Acad Mark Sci 43:115–135. https://doi.org/10.1007/s11747-014-0403-8

A Comprehensive Review of Innovation Adoption Types and Their Effects on the Project Performance

Majid G. Al-Nabae and Dania Sammani

Abstract In the current dynamic markets, characterized by competitive economies, innovation is deemed to be the peak of success in most business projects. However, studies dedicated to the relationship between innovation adoption and project performance, particularly the former's impact on the latter, still remains lacking. Thus, this paper's main objective is to shed light on the impact of innovation adoption on project performance through a thorough systematic review of literature dedicated to the subject. The literature review findings provide insight into the significance of innovation adoption in the context of project management. The study findings also contribute to the promotion of awareness concerning the innovation adoption role in enhancing project performance. Managers can use the findings to enhance project innovation capability through the acknowledgement of the different aspects of innovation.

Keywords Innovation · Innovation adoption · Project performance

1 Introduction

Innovation is a core component of the strategies of the organization and is a required resource for the organization to adopt in the business environment and to conduct effective management in the dynamic current markets [1]. In addition, innovation is a significant part of competitiveness in that it is integrated in services, products, processes and structure of the organization [2]. Several challenges exist in maintaining business competitiveness by leveraging opportunities and these include the ongoing technology enhancement and globalization, leading to the creation of new market forces, urging for the need to adopt innovation [3]. According to Midler et al. [4] innovation has become a vital issue in projects within organizations in the face of today's uncertainty, high competitiveness and technological changes.

M. G. Al-Nabae · D. Sammani (✉)
Faculty of Arts and Science, International University of Malaya-Wales, Kuala Lumpur, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_65

Moreover, innovation abilities development paves the way for introducing and enhancing new products and services and for the use of technologies and knowledge [5]. Based on the innovation theories that showed high promise in high-technology sectors, the process of manufacturing and innovation involves distinct individual actions [6].

Therefore, this study primarily aims to provide insight into the innovation types in project management by conducted a systematic review of literature dedicated to the subject. The first study contribution to literature is the presentation of a discussion of the innovation definitions and the types mentioned in literature. The second contribution to literature by seeding light on the way innovation adoption is related with project performance. Accordingly, the paper is organized in the following way; research methodology part of the paper presents the employed research method, the literature review part presents a thorough review of literature concerning innovation and project performance and the findings part enumerates the main findings of the review. Lastly, the paper provides the conclusions, limitations, and future developments under the conclusion part.

2 Research Methodology

The study takes account the review of related literature dedicated to project performance and innovation adoption to achieve the study objectives. The review covers innovation definition, types and its effect on projects and organizational performance. The research basically focuses on explaining the innovation adoption-project performance relationship.

The study adopted a three-step process of systematic literature review from Tranfield et al. [7] The initial step involves the identification of the keywords, which in the context of the study, are; innovation, innovation adoption and project performance. The selection of articles was based on the combination of the main keywords from the articles, keywords or abstract. The selected articles were all in English, published from January 2014 to June 2021. Moreover, the study only chose the articles in the Impact Factor and Scopus indexed journals to make sure that the articles are of high quality. As recommended by Tranfield et al. [7], the articles were reviewed to ensure that they are relevant to the study, and in this process, some articles were dropped from further analysis because although they were about project performance and innovation adoption types, they did not elaborate on project performance, while others were only limited themselves to project performance.

3 Literature Review

3.1 Definition of Innovation

Innovation refers to the process of producing new ideas and implementing them to generate or create new products, processes or services, which in turn could lead to the active domestic economic development, mitigation of unemployment and generation of profit for innovative projects [8]. Makkonen et al. [9] described innovation as one of the many significant elements that make up organizational competitiveness and is the effective adoption of new technology developments.

Meanwhile, Shaziet al. [10] referred to innovation adoption as a process comprising several connected phases that begins with generation of ideas and culminates with the implementation of services and products. Literature provides various definitions of the term innovation based on the field or context (e.g., government, academia, business), leading to misunderstandings and inconsistent interpretation of the meaning of the term, with some researchers using the term synonymously with invention, while others refer to it as something new, improved, new technology, and the like [11].

3.2 Innovation Types

Innovation has been categorized into different types in literature, with the common major types being technological and non-technological innovation and managerial innovation [12–15]. Specifically, technological innovation is the relationship between components, techniques, methods and processes that are required for product/service generation or introduction [8]. According to Zaman and Sedera [16] technological innovation plays a key role in responding to challenges of sustainability and thus, companies are turning into innovation to sustain their business activities (Fig. 1).

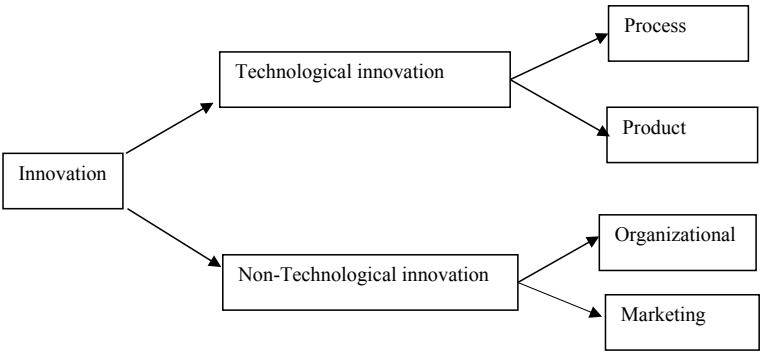


Fig. 1 Types of innovation

In literature non-technological innovation and its relationships have been extensively examined. For instance, Cerne [17] described non-technological innovation as the innovation pertaining to administration, organization and management, which is why it is often called administrative and organizational innovations and conceptualized opposite to that of technological innovation, which has its basis on innovation relating to process and product—rather non-technological innovation is related to new administrative systems, management activities and organizational structures [18]. More importantly, non-technological innovation has not been dwelled upon like its technological counterpart and was often overlooked as a determinant of performance through its critical effects on the same that in turn, effects the flexibility, productivity, quality and lead times of businesses [17]. While innovation concerns the organizational processes, it can also be about the organization itself or the technology that it adopts [19].

Related studies in literature including; Medrano and Olarte-pascual [20], Mensah and Acquah [21], Tan et al. [3] were of the consensus regarding the two main types of innovation, divided further into four classes (i.e., technological innovation that covers process and products, and non-technological innovation that covers marketing and organizational innovation).

3.3 Why is Innovation Important?

Innovation has been extensively acknowledged as an economic development and growth driver [22] and it can be a major competitive source of organization that is linked directly to the science fields of management and economics [23].

More importantly, organizations driven by the innovation strategy are focused on achieving growth performance, financial performance, learning and internal business processes development [24]. In fact, one of the critical factors of organizational competitiveness is the effective new technological innovation adoption [9]. In addition, innovations affect products quality, financial performance as well as operational performance and as such, new innovative services and products are deemed to be the key to competitive advantage sustainability [25].

In the realm of international business development and growth and global economic evolution, knowledge, technology, and innovation are three interconnected concepts and phenomena that are the major drivers, particularly in the past 50 years [26]. Along a similar Izmailov et al. [27] laid stress on organizational innovation adoption to facilitate an appropriate workplace environment and to promote technological innovation adoption process for the improvement of substantial performance. Also, several barriers to innovation investments were highlighted in Pellegrino and Savona's [28] study and these included technology, financial issues or the perception of adverse conditions. Therefore, large-sized firms have been increasingly focusing on their open innovations involving giving others access to their underused knowledge and assets and obtaining external inputs from groups for the purpose of R&D productivity [29].

Lastly, various schemes with a set of stages are related to decisions pertaining to the achievement of technological innovation adoption [30], and this explains why medium and small projects that had previous experiences in innovation have a higher likelihood towards achieving work success and generating profitability compared to their non-innovating counterparts [31].

3.4 Innovation and Adoption and Project Performance

The innovation concept is core to new industries, competitive advantages and firm performance development as well as management effectiveness of public and business firms [18]. Some prior studies advocated the importance of innovation as a determinant of organizational performance e.g., (Karabulut [24]; Mensah and Acquah [21], Sheshi and Kercini [32]). For this reason, project management practices that have product and process innovation are used by enterprises for competitive advantage sustainability [33]. Moreover, open innovation initiatives are being developed in some companies to enhance the navigation of projects in a complex marketplace [34]. In fact, industrial projects are into innovation development to ensure successful outcomes [35].

Furthermore, agile project management (APM) methods have been noted to be extensively used in organizations for innovation development [36]. These methods have become essential for both high-tech and SMEs, whereby their performances greatly differ owing to the different levels of their capability for innovation and the dynamics within their internal and external environments [37].

In this regard, the project success when it comes to innovations is affected by the project team's coordination, leadership and commitment. With regards to product innovation, interactions occur frequently between R&D department, teams and other functions for effective project outcome [38]. In the context of Western economies, innovation and technological leadership form the key to competitiveness, particularly in manufacturing industries for project performance enhancement [39]. Therefore, it can be stated that project management innovation redefines the faced challenges rather than furnishing solutions to the problem—through this manner, it promotes creative thinking and innovative ideas [40]. In the same line of study, innovation adoption process was considered to cover reconfiguration tools of resources base and performance improvement for sustainable competition (Makkonen et al. [9]). Also, lack of innovation and creativity may lead to lack of competitiveness [41] and in this regard, innovation project performance depends on organizational features [8].

Similarly, other studies in literature claimed that innovation is a strategic method that companies can adopt for competitive advantage and for international market survival [24]. Innovation in fact has a significant impact on business performance and the relationship between the two variables largely hinges on different factors (e.g., trade openness, industry structure, capital markets, labor markets, copyrights and property, educational system, governmental policies, national culture etc.) [32].

Along the same caliber of study, Hamidi and Benabdeljlil [42] revealed that innovation has a critical role in value generation and in performance enhancements of organizations. Therefore, both creativity and innovation has to be stressed in firms for performance enhancements, expertise maximization, ability and motivation enhancement for overcoming challenges, and creative thinking development in order to come up with new and innovative business ideas [41]. Finally, innovation can be one of the mechanisms and strategies utilized to enhance the performance of the project {Formatting Citation}. but only product innovation was revealed by Geldes et al. [14] to influence performance throughout the industries.

4 Discussion and Findings

Under this section, the paper presents the analysis results and explains the way innovation adoption leads to successful projects performance.

4.1 Innovation Types

As mentioned, innovation has been categorized into two major divisions in literature and they are technological and non-technological innovation/managerial innovation (e.g., [12–14]. Meanwhile, other studies in literature including Acosta et al. [11], Medrano and Olarte-pascual [20], Mensah and Acquah [21], and Tan et al. [3] categorized technological innovation into process and product innovation, and non-technological innovation into marketing and organizational innovation, theorizing their linkage with the performance of the organization.

4.2 Innovation Adoption and Project Performance

From the literature reviewed, the paper found prior studies, findings to support the premise that innovation adoption strategy is crucial for the performance of projects, with a positive innovation-organizational performance relationship—in other words, innovation has a key role in organizational performance improvement (e.g., [36, 38–40].

In sum, innovation adoption in project management appear to be lauded by prior studies to contribute to competitiveness [2] to enhance the success of projects [38], produce value to the firms [42], enhance the performance of the project [39], drive growth of the economy [22] add to project management and research value [33], and create innovative solutions [40].

5 Limitations

The systematic review conducted by the present paper has limitations, the most important of which is the emphasis on high-quality research while choosing relevant and peer-reviewed academic articles concerning innovation adoption and project performance from selected databases and top journals. Relevant publications may have been neglected by following this strategy.

6 Future Research Directions

In this paper, only innovation types and their effects on project performance have been mentioned as a result of the reviewed relevant literature. This leaves plenty of avenues for future research to follow; first, majority of reviewed empirical studies employed quantitative survey designs. Future studies can thus adopt qualitative research designs/case studies to minimize the gap on literature concerning innovation adoption and implementation for project management. In this regard, case studies can provide more insight into the role of various innovation types as they are the core to innovation adoption development among many projects. Also, majority of the reviewed studies focused on the level of the organization, necessitating the focus on project types in future studies.

Another avenue for future studies is the examination of the relationship between project performance and specific types of innovation (e.g., product, process, organization and marketing innovation). This literature gap calls for minimizing through further examination following this research avenue, where future studies can examine other innovation types that may need concentration in project completion.

7 Implications for Research and Practice

The present literature review successfully synthesized research findings to highlight meta-level evidence and other areas needed to be extensively researched—this constitutes a critical element of theoretical frameworks and conceptual models development. More specifically, three primary innovation adoption types themes were found by the review in the context of project management. The first theme is the conceptualization of innovation adoption either as a process or outcome, influencing project performance, and the second theme is the project context, where innovation adoption has been frequently conceptualized as a process that improves the project performance and the third theme, is the premise that innovation adoption is frequently examined in the organization or in the technological sector—thus, this study essentially contributes to literature on the innovation adoption effects on project performance and innovation types.

Based on the study findings, innovation adoption's relation to performance could lay the basis for firms and projects to obtain competitive advantage sustainability. Moreover, the four innovation types' conceptualizations, namely process innovation, product innovation, organizational innovation and marketing innovation have also been established in the context of project management. In this way, project management practitioners can make use of the findings to promote different innovation types for enhanced performance.

8 Conclusion

This research provided the findings on the effect of innovation adoption on project performance obtained from a thorough and systematic review of prior literature on innovation, its types and its effects on project performance. The review focused on the way innovation adoption influences project performance while at the same time highlighting its role as a vital strategy to promote the enhanced performance of projects in a globalized every-changing, dynamic market. Existing literature indicate the inevitability of adopting innovation in today's world and a must for enhancing performance, owing to its significant influence on project performance. Innovation investments are essential, particularly when it comes to enhancing performance outcomes among projects.

To conclude, it is stated that innovation has been extensively acknowledged as the pinnacle of success factors in the project performance context, assisting in solving problems and critical-thinking among its adopters, and increasing the collective capacity for creative imagination and diversity in organizational practices. Evidently, innovation adoption in project performance lays the basis for successful projects and sustainable competitive advantage of firms.

References

1. Hanif MI, Akhtar MS, Nawaz MK, Abbas SJ (2017) Impact of innovation generation and adoption on firm's performance. South Asian J Banking Soc Sci 3(1)
2. Sousa P, Tereso A, Alves A, Gomes L (2018) Implementation of project management and lean production practices in a SME Portuguese innovation company. Procedia Comput Sci 138:867–874
3. Tan J, Cao H, Kong X (2019) Do major customers promote firms' innovation? China J Account Res 12(2):209–229
4. Midler C, Killen CP, Kock A (2016) Project and innovation management: bridging contemporary trends in theory and practice. Project Manage J 47(2,3–7):269–272
5. Carpio CR, Figueroa WE, Alvarado MP (2015) Innovation management system of ecuador. Procedia-Soc Behav Sci 157–166
6. Hogan SJ, Coote LV (2014) Organizational culture, innovation, and performance: a test of Schein's model. J Bus Res 67(8):1609–1621

7. Tranfield D, Denyer D, Smart P (2003) Towards a methodology for developing evidence-informed management knowledge by means of systematic review. Br J Manag 14:207–222
8. Kogabayev T, Maziliauskas A (2017) The definition and classification of innovation. Holistica 8(1)
9. Makkonen H, Johnston WJ, Javalgi RRG (2016) A behavioral approach to organizational innovation adoption. J Bus Res 69(7):2480–2489
10. Shazi R, Gillespie N, Steen J (2015) Trust as a predictor of innovation network ties in project teams. Int J Project Manage 33(1):81–91
11. Acosta B, Acosta M, Espinoza B (2016) Understanding innovation based on company optics: interpretation mistakes on the types of innovation developed. RAI Revista de Administração e Inovação 13(4):295–304
12. Anzola-román P, Bayona-sáez C, García-marco T (2018) Organizational innovation, internal R&D and externally sourced innovation practices: Effects on technological innovation outcomes. J Bus Res 91(June):233–247
13. Damanpour F, Sanchez-Henriquez F, Chiu HH (2018) Internal and external sources and the adoption of innovations in organizations. Br J Manag 00:1–19
14. Geldes C, Felzensztein C, Palacios-fenech J (2016) Industrial marketing management technological and non-technological innovations, performance and propensity to innovate across industries: The case of an emerging economy. Ind Mark Manage 61:55–66
15. Henry D, Walker T, Harley J, Mills A (2015) Performance of project alliancing in Australasia: a digest of infrastructure development from 2008 to 2013. Constr Econ Build 15(1):1–18
16. Zaman B, Sedera D (2015) Green information technology as administrative innovation—organizational factors for successful implementation: literature review. australasian conference on information systems 2015, Adelaide, Loeser 2013, 12
17. Cerne M (2016) Non-technological innovation research: evaluating the intellectual structure and prospects of an emerging fi eld. Scand J Manag 32:69–85
18. Damanpour F (2014) Footnotes to research on management innovation. Organ Stud 35(9): 1265–1285
19. Loi R, Lam LW, Ngo HY, Cheong S (2015) The differentiated impacts of organizational innovation practices on technological innovation persistence. Eur J Innov Manag 18(1):110–127
20. Medrano N, Olarte-pascual C (2016) An empirical approach to marketing innovation in small and medium retailers: an application to the Spanish sector. Contemp Econ 10(3):205–216
21. Mensah FB, Acquah ISK (2015) The effect of innovation types on the performance of small and medium D sized and enterprises and in the Sekondi—Takoradi Metropolis. Arch Bus Res 3(3):77–98
22. Medase SK, Abdul-Basit S (2020) External knowledge modes and firm-level innovation performance: empirical evidence from sub-Saharan Africa. J Innov Knowl 5(2):81–95
23. Taques FH, López MG, Basso LF, Areal N (2021) Indicators used to measure service innovation and manufacturing innovation. J Innov Knowl 6(1):11–26
24. Karabulut AT (2015) Effects of innovation types on performance of manufacturing firms in Turkey. Procedia Soc Behav Sci 195:1355–1364
25. Kafetzopoulos D, Psomas E (2015) The impact of innovation capability on the performance of manufacturing companies the Greek case. J Manuf Technol Manag 26(1):104–130
26. Andersson A, Ciabuschi F (2017) Organizational innovation, technological innovation, and export performance: the effects of innovation radicalness and extensiveness. Int Bus Rev 26(2):324–336
27. Izmailov A, Korneva D, Kozhemiakin A (2016) Effective project management with theory of constraints. Procedia Soc Behav Sci 229:96–103
28. Pellegrino G, Savona M (2017) No money, no honey? Financial versus knowledge and demand constraints on innovation. Res Policy 2016:1–12
29. Brunswicker S, Chesbrough H (2018) The adoption of open innovation in large firms. Res Technol Manage 61(1):35–45

30. Miranda MQ, Farias JS, Schwartz C de A, Almeida JPL de (2016) Technology adoption in diffusion of innovations perspective: introduction of an ERP system in a non-profit organization. RAI Revista de Administração e Inovação 13(1):48–57
31. Love JH, Roper S (2015) SME innovation, exporting and growth: a review of existing evidence. Int Small Bus J Res Entrepreneurship 33(1):28–48
32. Sheshi A, Kercini D (2017) The role of technological, technical and administrative innovation in the performance of the Sme'S in Albania. European J Bus, Econ Accountancy 5(2):95–103 (2017)
33. Severo EA, Sbardelotto B, de Guimarães JCF, de Vasconcelos CRM (2020) Project management and innovation practices: backgrounds of the sustainable competitive advantage in Southern Brazil enterprises. Prod Plann Control 31(15):1276–1290
34. Barbosa APPL, Salerno MS, Brasil VC, de Nascimento PTS (2020) Coordination approaches to foster open innovation R&D projects performance. J Eng Technol Manage JET-M, 58(October)
35. Saunila M (2020) Innovation capability in SMEs: a systematic review of the literature. J Innov Knowl 5(4):260–265
36. Lill PA, Wald A, Gleich R (2020) Agility and the role of project—internal control systems for innovation project performance. Int J Innov Manag 24(7):1–29
37. Ju X, Ferreira FAF, Wang M (2020) Innovation, agile project management and firm performance in a public sector-dominated economy: empirical evidence from high-tech small and medium-sized enterprises in China. Socio-Econ Plann Sci 72:100779
38. Gemünden HG, Lehner P, Kock A (2018) The project-oriented organization and its contribution to innovation. Int J Project Manage 36(1):147–160
39. Schoper YG, Wald A, Ingason HT, Fridgeirsson TV (2018) Projectification in western economies: a comparative study of Germany, Norway and Iceland. Int J Project Manage 36(1):71–82
40. Sanz-Llopis J, Ostermann M (2020) Innovation in project management through framing and challenge redefinition. Int J Manag Proj Bus 13(4):745–766
41. Dessie WM, Ademe AS (2017) Training for creativity and innovation in small enterprises in Ethiopia. Int J Training Dev 21(3):224–234
42. Zuhairy M, Tajuddin M, Iberahim H, Ismail N (2015) Relationship between innovation and organizational performance in construction industry in Malaysia. Univ J Ind Bus Manage 3(4):87–99

The Role of Digital Platforms in Alleviating the Fear of Entrepreneurial Failure: A New Conceptual Framework

Khalid Mady, Mohamed Elsotouhy, and Mohamed Battour

Abstract The literature on digital entrepreneurship has paid extensive attention to investigating the role of digital technologies in the process of entrepreneurship based on socio-technical perspectives. The role of digitalization on psychological aspects of entrepreneurship has been completely ignored despite these psychological aspects such as fear of failure, anxiety, and lack of confidence act as a deterrent for nascent entrepreneurs from starting and scaling up their ventures. This paper, hence, aims to conceptualize the role of digital platforms in mitigating the fear of failure in the idea-to-venture process. Two proposed paths are conceptualized through which engaging in digital platforms can help nascent entrepreneurs alleviate their fear of failure. This conceptualization will be appreciated by academic researchers in the entrepreneurship area, entrepreneurs, and digital platform providers.

Keywords Digital entrepreneurship · Digital platforms · Fear of failure · Crowdsourcing platforms

1 Introduction

Entrepreneurship has recently been conceived as an engine of global economic and societal growth [1]. Let alone the employment opportunities that entrepreneurs provide both for themselves and for the others, entrepreneurial activities also play a

K. Mady (✉)
Faculty of Business and Economics and social development, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia
e-mail: Khaled.Mady@com.kfs.edu.eg

M. Elsotouhy
Faculty of Commerce, Mansoura University, Mansoura, Egypt

M. Battour
College of Business Administration, University of Sharjah, Sharjah, UAE
e-mail: mbattour@sharjah.ac.ae

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_66

vital role in enhancing economic performance in several ways. First, entrepreneurs often use new products, advanced technologies or innovative production processes to penetrate the markets. Additionally, productivity and competition benefit from entrepreneurs entering the market. Latterly, structural transformation has become accelerated thanks to entrepreneurial orientation whereby it is expected that economies will stagnate without new entrepreneurs [2, 3]. By transforming into 'sharing-economy platforms,' entrepreneurs have faced more challenges regarding how to utilize digital platforms to facilitate entrepreneurial activity [4]. Therefore, many previous studies have addressed the role of digital platforms in facilitating entrepreneurship processes [5, 6].

More than half of the population in 36 of 50 economies have the required skills, experience and knowledge to initiate their own business, while in 42 of the 50 economies fewer than half of those who recognise commercial opportunities would be deterred by a fear of failure [7]. Hence, psychological factors such as a fair of failure and lack of self-confidence may lead nascent entrepreneurs to forego their business opportunity [8]. The literature concerning the fear of entrepreneurial failure concentrates increasingly on two main research streams: the impact of these feelings on the entrepreneurial intention and entrepreneurial activity [8–10] and the causes and sources of these feelings. Concerning the latter direction of research, Kollmann et al. [11] demonstrate that fear of failure is a result of perceiving various obstacles related to resources, the market and social capital and it is subsequently detrimental to entrepreneurs' ability to evaluate and exploit opportunities. In the same vein, Cacciotti et al. [12] grouped the sources of fear of failure depending on the perceived concerns of nascent entrepreneurs into seven main clusters such as "personal ability, the potential of the idea, and the venture's ability to execute". However, the question of how nascent entrepreneurs can mitigate the fear of entrepreneurial failure is still in urgent need of being addressed.

Digital technologies have recently become the most critical contributor to the creation and conception of new business ventures [13, 14]. Platformization has placed collaborative and collective intelligence at the core of effective and sustainable business initiatives [15, 16]. Online platforms as purposive technology whereby businesses are able to connect and interact can be beneficial for smaller businesses to establish, innovate and scale-up quickly [6]. Open platforms such as Kickstarter, InnoCentive, and F6S can help entrepreneurs by providing finance, resources and support to make entrepreneurial ideas a reality and foster entrepreneurship [5, 6]. Consequently, these technologies can reduce the barriers and obstacles that face nascent entrepreneurs across the idea-to-venture phases and may, in turn, help to alleviate their fear of failure. The objective of the study is to conceptualize the role of online platforms in reducing the barriers that lead to a fear of failure for entrepreneurs as one of the elements of digital technologies. The reason for such a conceptualization is to help entrepreneurs utilize the properties and services on these platforms to mitigate their fear of entrepreneurial failure. Developing and testing this conceptualization will be of interest to several parties including future researchers, entrepreneurs, and digital platform providers.

2 Literature Review

2.1 Fear of Failure: Definitions

Fear of failure is conceived of as an intrinsic feature of the entrepreneurs' experience [17]. Entrepreneurs on their way to starting and establishing their new venture face the fear which is related to various tasks that may lead to failure such as opportunity exploitation and idea promotion. There are two main different streams of research to define the fear of failure concerning entrepreneurship [18]. Some studies define the concept of fear of failure as an unchangeable 'fixed' trait affecting social incentives given to entrepreneurs [19, 20]. Thus, fear of failure makes individuals avoid situations that involve a potential risk. By contrast, other studies define it as a negative emotional state depending on the cognitive appraisals of the surrounding environment and uncertainty about potential opportunities [21, 22]. The current study adopts the latter perspective whereby the fear of failure tends to be adapted and alleviated by following strategies and making efforts to reduce uncertainty.

2.2 Fear of Failure: Sources

According to appraisal theory, fear is a temporary emotional state which is extracted from individuals' appraisals of actual or anticipated events [23, 24]. A fear of failure can be experienced when individuals recognize the anticipated threats to their ability to achieve a certain meaningful goal [25]. Consequently, entrepreneurs' fear of failing can be triggered by the obstacles threatening the founding and success of a new venture. Prior studies articulate that these fears facing nascent entrepreneurs can be divided into two types: exogenous fear of failure which refers to fears related to obstacles external to entrepreneurs [11] and endogenous fear of failure which refers to fears related to obstacles internal to entrepreneurs [17].

- **Internal sources of fear of entrepreneurial failure**

1. Idea or opportunity-related fear: Finding a good idea and exploiting opportunities are the crucial issues in entrepreneurship actions. Studies have revealed that the inability to find potential opportunities and their lack of quality are significant sources of the failure of ventures [12].
2. Self-efficacy-related fear (lack of competency fear): Self-efficacy is regarded as one of the basic determinants of success in entrepreneurship [26]. Competence consists of knowledge, abilities and attitudes which allow individuals to be effective in the workplace and their everyday life [27].

- **External sources of fear of entrepreneurial failure**

1. Resources-related fear: A lack of access to resources is considered to be one of the sources of entrepreneurship failure [28, 29].
2. Social network-related fear: A lacking of social esteem can be represented as a source of the fear of failure in entrepreneurship [12]. In the entrepreneurial process there are multiple stakeholders including investors, business partners, customers, family and employees who need to be satisfied by the entrepreneur [12].
3. Market-related fear: The success of new a venture depends on its ability to cope with the external environment such as marketing and competition challenges. Failing to market a venture's products or services constitutes a serious threat arousing the fear that might deter nascent entrepreneurs [12].

2.3 Digital Platforms and Entrepreneurial Processes

Business and social activities are nowadays facilitated by platforms in what is called the 'platform economy'. A digital platform is an internet-based technology in which businesses or individuals can be connected and share data, knowledge, and experiences [6, 14]. These platforms can be categorized into four groups [30]. First, transaction platforms are designed to facilitate commercial and social interactions between various users such as suppliers and buyers. Amazon, eBay and Uber are examples of transaction platforms through which goods and service can be exchanged. Second, innovation platforms provide a foundation upon which developers or innovators can get started and develop their idea, product or service. Third, an integrated platform can bring together the capabilities of both transaction platforms and innovation platforms. A case in point is Amazon which besides serving as a transaction platform, also allows independent sellers to develop and offer their products on its platform. Finally, investment platforms act as either holding companies that run a portfolio of other platforms or active platform investors.

Many digital platforms are designed to target new ventures and nascent entrepreneurs during idea-venture processes. Entrepreneurs that utilise digital platforms have a promising opportunity to develop their ideas and transform them into successful ventures [29]. Most of these platforms increasingly rely on the idea of crowdsourcing to connect entrepreneurs with communities of experts and innovators with various backgrounds to solve real business challenges [31]. For example, IBM Innovation Jam and InnoCentive tap into the collective thinking of their crowdsourcing to remove the barriers that face entrepreneurs during their entrepreneurial journey. Besides, these digital platforms provide entrepreneurs with resources such as customer-related knowledge. Crowdfunding platforms introduce an alternative form of finance to eliminate any financial barriers facing ventures [32, 33]. Even these

digital platforms that don't target entrepreneurs directly can also benefit by providing their infrastructure and services for entrepreneurs such as Facebook and Uber.

3 Methodology

The current study is a conceptual paper that can lead to two main theoretical contributions [34]. First, a conceptual paper can provide new, alternative or emerging perspectives based on the prior literature [34, 35]. The study addresses the role of digital platforms in the entrepreneurship process but from a divergent perspective which is the psychological perspective. Second, this type of paper proposes a conceptual model taking into consideration the boundary conditions and the potential mechanisms that can explain a particular phenomenon [34, 36]. The study proposes a new model conceptualizing the relationship between numerous constructs including fear of failure, entrepreneurship, and digital platforms. Moreover, this conceptual model provides the mechanisms explaining how digital platforms help nascent entrepreneurs to alleviate their fear of failure. Conceptual papers often raise a number of propositions that should be testable, although testing these propositions doesn't fall within the scope of this type of research.

4 Proposed Conceptual Framework

This study depends on the conceptualization that the fear of failure is largely dependent on the perception that entrepreneurs are adversely affected by these potential threats in the idea-to-venture processes. Apart from avoidance-based strategies, entrepreneurs have a powerful choice regarding the fear of failure by meeting these threats [37]. This study conceptualizes that digital platforms could have a dual role in alleviating the fear of failure through two main paths (see Fig. 1):

4.1 Surmounting the Obstacles Facing Entrepreneurs

External obstacles can arouse the fear of failure that makes entrepreneurs devalue their promising opportunities and, in turn, forsake their business ideas [11]. Access to resources (e.g. finance, knowledge and capabilities) is one of the most important functionalities of digital platforms such as Kickstarter and F6S [38, 39]. Crowdfunding platforms offer an alternative source of funding streams which allow entrepreneurs and start-ups to fund their ideas and ventures [16]. These platforms operate a risk- and profit-sharing approach that allows entrepreneurs to share their venture risks [33]. Digital innovation platforms such as WhataVenture and GEN

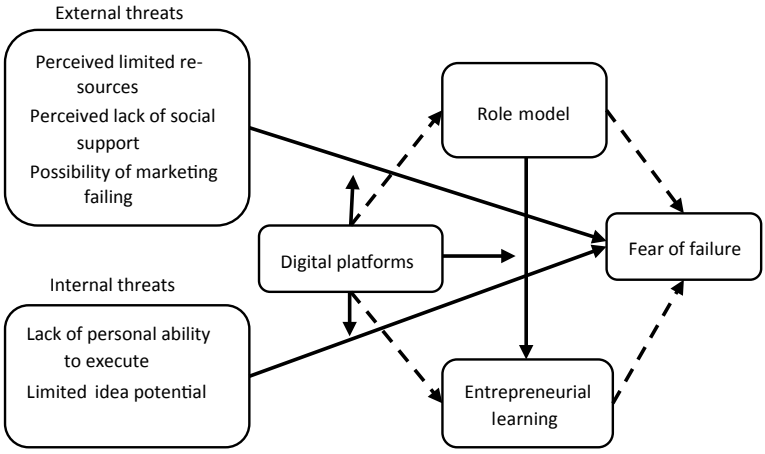


Fig. 1 Conceptual framework

can also contribute by providing entrepreneurs with critical knowledge and capabilities, allowing them to conceptualize and convert their ideas into new ventures [5, 38].

Digital platforms such as Amazon Marketplace and Facebook Marketplace provide the digital infrastructure for small businesses to offer their products and services to either local or global markets [16, 40]. Also, entrepreneurs with a fear of marketing failure can tap into the data obtained in social media platforms to obtain detailed insight into market demand, customers' views and feedback regarding certain brands or products [41]. As such, some digital innovation platforms use social analytics and data mining tools to analyze social media content and provide meaningful insights that can help start-ups to understand business and market trends [42]. Participation in various digital platforms provides social interaction through which entrepreneurs can gain social support from peers [31].

Proposition 1 *Engaging in digital platforms may mitigate external threats (perceived limited resources, venture's inability to market, and perceived lack of social support) and, in turn, alleviate the fear of entrepreneurial failure.*

The fear of entrepreneurial failure can also be triggered by a perception of internal personal or venture threats which can be related to a lack of personal ability to execute the stages of the idea-to-venture process and lower the potential of the entrepreneurial ideas [12]. Digital platforms have played distinctive roles as entrepreneurship incubators or accelerators [43, 44]. Regarding digital platforms, communities and collective intelligence can contribute to different stages of entrepreneurship by sharing knowledge, providing resources, and connecting with experts and problem solvers [5]. However, the ideation process has received greater attention from crowdsourcing platforms [31]. Open and active discussion among community members and peers on crowdsourcing platforms helps entrepreneurs to

collect and manage ideas as well as guiding their ideas so that they are successful in introducing solutions to potential problems [43]

Proposition 2 *Engaging in digital platforms may mitigate internal threats (limited idea potential, and lack of personal capabilities) and, in turn, alleviate the fear of entrepreneurial failure.*

4.2 Empowering Entrepreneurs to Face the Fear of Failure

Role models

A role model is an inspirational person who can be imitated by others [45]. For entrepreneurs, a role model represents a driving factor when starting their new ventures [10, 45]. As concluded by [46], the more role models entrepreneurs have observed, the more likely they are to have close ties with their entrepreneurial stages. Importantly, observing successful entrepreneurs can also help to reduce the fear of failure and, therefore, plays an essential role in entrepreneurship activities [10, 37]. Entrepreneurial role models can help to alleviate perceived fears of business failure by tapping into the knowledge and legitimacy acquired from observing and contacting with other entrepreneurs [47]. Digital platforms have intentionally presented success stories as examples that can be emulated by aspiring entrepreneurs. Networks of experts or other peers on these platforms can constitute role models for nascent entrepreneurs.

Proposition 3 *Digital platforms, in particular the innovation platforms involving a network of experts and peers from different backgrounds, can provide role models that help nascent entrepreneurs to overcome the fear of entrepreneurial failure.*

Entrepreneurial learning

Entrepreneurial learning can help to ensure that entrepreneurs are undeterred by the fear of failing [48]. Entrepreneurial learning is defined as a continuous and experiential process through which entrepreneurs develop the knowledge necessary to be effective in establishing and managing new ventures [49]. Entrepreneurial learning can broaden entrepreneurs' understanding of the resources needed for entrepreneurship and the challenges that may appear in the idea-to-venture process which can alleviate the anxiety and fear of entrepreneurial failing [10]. Digital platforms accelerate the sharing, acquisition and creation of knowledge, thus developing opportunities for new skills that can be learned and for innovative designs that can be created [50]. A crowdsourcing platform is perceived as a dynamic knowledge alignment system enabling entrepreneurs to learn from the heterogeneous knowledge available from various members of a crowd such as customers and other entrepreneurs [51, 52]. To provide opportunities for learning and knowledge sharing, digital platforms such as Alibaba organize a series of

training courses for different types of entrepreneurs [53]. Even social networks provide important platforms for entrepreneurship learning by helping to share information and mitigate the problem of information failure [54].

Proposition 4 *Digital platforms, especially innovation platforms involving a network of experts and peers from different backgrounds, can play a crucial role in entrepreneurial learning to alleviate the fear of entrepreneurial failure.*

Besides motivation and inspiration, entrepreneurial role models can be conceived of as a source of learning for potential entrepreneurs [47]. As per social learning theory, observing and interacting with others, particularly those regarded as entrepreneurial role models, can catalyze entrepreneurial learning by providing opportunities to gain an in-depth understanding of the tasks and knowledge required to participate in entrepreneurial activities successfully [47]. As for nascent entrepreneurs, role models can constitute a source of indirect experience and also enhance the potential of entrepreneurs to discover and exploit business opportunities [10]. Entrepreneurial learning outcomes (about oneself and about business) do not depend solely on the role models observed but rather they are contingent on social contexts (e.g. home, education, work and business) for the pre-and post-start-up stages [46]. The current study attempts to reconceptualize the influence of role models in entrepreneurial learning but in a different social context, namely, digital platforms such as innovation platforms and crowdsourcing platforms.

Proposition 5 *Digital platforms, especially innovation platforms involving a network of experts and peers with different backgrounds, provide nascent entrepreneurs with role models (or success stories) that can enhance entrepreneurial learning.*

5 Conclusion and Expected Implications

The empirical literature concerning digital entrepreneurship focuses its attention on the main question of the role of digital technologies in the entrepreneurial process, drawing on socio-technical perspectives [13, 44, 55]. However, the psychological aspect of digitalization and its role in the idea-to-venture process is completely ignored. Based on the psychological perspective, this paper aims to conceptualize the significance of digital platforms in mitigating the fear of entrepreneurial failure, which has been regarded as a destructive psychological feeling preventing nascent entrepreneurs from starting new ventures. This conceptualization assumes that using different types of digital platforms could have beneficial side-effects by mitigating such fear. Digital platforms can first help entrepreneurs to mitigate this fear by surmounting the obstacles or threats facing entrepreneurs. Based on the context of threats, the conceptual framework depends on two types of threats, namely, external threats and internal threats. Secondly, digital platforms can also

provide nascent entrepreneurs with entrepreneurial role models and entrepreneurial learning capabilities that empower them to face the fear of failure.

The study has broadened the knowledge on two streams of literature, namely, digital entrepreneurship and the fear of entrepreneurial failure by proposing a conceptual framework and five research propositions. However, the validation of these propositions will depend on empirical studies that will use the proposed model. The results of examining the proposed conceptual framework could have serious implications for future research, entrepreneurs and digital platform providers. Testing the framework conceptualized here would extend our knowledge of the role of digital technologies in entrepreneurship from a psychological perspective. This paper can hence serve as the basis for future research on the impact of digitalization on psychological factors such as the fear of failure and lack of self-confidence. Entrepreneurs can deal with their fear of failure by utilising a digital platform that can help them. Another potential practical implication is that digital platform providers can benefit from examining this conceptual framework to develop and offer features and properties that consider the entrepreneurs' fear of failure.

References

1. Audretsch DB, Keilbach MC, Lehmann EE (2006) Entrepreneurship and economic growth. Oxford University Press
2. Kritikos AS (2015) Entrepreneurship and economic growth. Int Encycl Soc Behav Sci Second Ed 675–680
3. Acs ZJ, Szerb L (2007) Entrepreneurship, economic growth and public policy. Small Bus Econ 28(2–3):109–122
4. Ditta-Apichai M, Kattiyapornpong U, Gretzel U (2020) Platform-mediated tourism micro-entrepreneurship: implications for community-based tourism in Thailand. J Hosp Tour Technol 11(2):223–240
5. Elia G, Margherita A, Passiante G (2020) Digital entrepreneurship ecosystem: how digital technologies and collective intelligence are reshaping the entrepreneurial process. Technol Forecast Soc Change 150(October 2019):119791
6. Hsieh YJ, Wu YJ (2019) Entrepreneurship through the platform strategy in the digital era: Insights and research opportunities. Comput Human Behav 95:315–323
7. Bosma N, Hill S, Ionescu-Somers A, Kelley D, Levie J, Tarnawa A (2020) Global entrepreneurship monitor 2019/2020 Global Report
8. Martins I, Monsalve JPP, Martinez AV (2018) Self-confidence and fear of failure among university students and their relationship with entrepreneurial orientation: evidence from Colombia. Acad Rev Latinoam Adm 31(3):471–485
9. Tsai KH, Chang HC, Peng CY (2016) Refining the linkage between perceived capability and entrepreneurial intention: roles of perceived opportunity, fear of failure, and gender. Int Entrep Manag J 12(4):1127–1145
10. Kong F, Zhao L, Tsai CH (2020) The relationship between entrepreneurial intention and action: the effects of fear of failure and role model. Front Psychol 11(March):1–9
11. Kollmann T, Stöckmann C, Kensbock JM (2017) Fear of failure as a mediator of the relationship between obstacles and nascent entrepreneurial activity—an experimental approach. J Bus Ventur 32(3):280–301

12. Cacciotti G, Hayton JC, Mitchell JR, Giazitzoglu A (2016) A reconceptualization of fear of failure in entrepreneurship. J Bus Ventur 31(3):302–325
13. von Briel F, Davidsson P, Recker J (2018) Digital technologies as external enablers of new venture creation in the it hardware sector. Entrep Theory Pract 42(1):47–69
14. Battour M, Salaheldeen M, Mady K (2021) Halal tourism: exploring innovative marketing opportunities for entrepreneurs. J Islam Mark (2021)
15. Anderson C (2014) Makers: the new industrial revolution. Crown Business
16. Nambisan S, Siegel D, Kenney M (2018) On open innovation, platforms, and entrepreneurship. Strateg Entrep J 12(3):354–368
17. Cacciotti G, Hayton JC, Mitchell JR, Allen DG (2020) Entrepreneurial fear of failure: scale development and validation. J Bus Ventur 35(5):106041
18. Cacciotti G (2015) Fear of failure in entrepreneurship: a review, reconceptualization and operationalization. The University Of Warwick
19. Vaillant Y, Lafuente E (2007) Do different institutional frameworks condition the influence of local fear of failure and entrepreneurial examples over entrepreneurial activity? Entrep Reg Dev 19(4):313–337
20. Hessels J, Grilo I, Thurik R, van der Zwan P (2011) Entrepreneurial exit and entrepreneurial engagement. J Evol Econ 21(3):447–471
21. Li Y (2011) Emotions and new venture judgment in China. Asia Pacific J Manag 28(2):277–298
22. Mitchell JR, Shepherd DA (2011) Afraid of opportunity: the effects of fear of failure on entrepreneurial action. Front Entrep Res 6(2):195–209
23. Roseman IJ, Smith CA (2001) Appraisal theory: overview, assumptions, varieties, controversies. In: Appraisal processes in emotion: theory, methods, research. Oxford University Press, New York, NY, US, pp 3–19
24. Scherer KR, Schorr A, Johnstone T (2001) Appraisal processes in emotion: theory, methods, research. Oxford University Press, New York, NY, US
25. Hayton JC, Cacciotti G (2014) Fear of failure and entrepreneurship: a review and direction for future research. Enterp Res Cent 24:1–62
26. Wennberg K, Pathak S, Autio E (2013) How culture moulds the effects of self-efficacy and fear of failure on entrepreneurship. Entrepreneurship and regional development, 25(9–10):756–780
27. Sartori R, Tacconi G, Caputo B (2015) Competence-based analysis of needs in VET teachers and trainers: an Italian experience. Eur J Train Dev 39(1):22–42
28. Fomina Y, Chahine T (2019) Social entrepreneurship factors of success and failure in the Omsk Region of Russia BT—responsible people: the role of the individual in CSR, entrepreneurship and management education. In: Farache F, Grigore G, Stancu A, McQueen D (eds) Cham: Springer International Publishing, pp 89–111
29. Nambisan S, Baron RA (2019) On the costs of digital entrepreneurship: role conflict, stress, and venture performance in digital platform-based ecosystems. J Bus Res (November 2018):0–1
30. Evans PC, Gawer A (2016) The rise of the platform enterprise a global survey. Cent Glob Enterp 1–30
31. Bakici T (2020) Comparison of crowdsourcing platforms from social-psychological and motivational perspectives. Int J Inf Manage 54:102121
32. Cavallo A, Ghezzi A, Dell'Era C, Pellizzoni E (2019) Fostering digital entrepreneurship from startup to scaleup: the role of venture capital funds and angel groups. Technol Forecast Soc Change 145(May):24–35
33. Richter C, Kraus S, Brem A, Durst S, Giselbrecht C (2017) Digital entrepreneurship: innovative business models for the sharing economy. Creat Innov Manag 26(3):300–310
34. Post C, Sarala R, Gatrell C, Prescott JE (2020) Advancing theory with review articles. J Manag Stud 57(2):351–376
35. Gilson LL, Goldberg CB (2015) Editors' comment: so, What is a conceptual paper? Gr Organ Manag 40(2):127–130

36. Cropanzano R (2009) Writing nonempirical articles for journal of management: general thoughts and suggestions. J Manage 35(6):1304–1311
37. Wyrwich M, Sternberg R, Stuetzer M (2019) Failing role models and the formation of fear of entrepreneurial failure: a study of regional peer effects in German regions. J Econ Geogr 19 (3):567–588
38. Nambisan S, Lyytinen K, Majchrzak A, Song M (2017) Digital innovation management: reinventing innovation management research in a digital world. MIS Q 41(1):223–238
39. Ngoasong MZ (2018) Digital entrepreneurship in a resource-scarce context: a focus on entrepreneurial digital competencies. J Small Bus Enterp Dev 25(3):483–500
40. Ayedh A, Echchabi A, Battour M, Omar M (2020) Malaysian Muslim investors' behaviour towards the blockchain-based Bitcoin cryptocurrency market. J Islam Mark
41. Olanrewaju AST, Hossain MA, Whiteside N, Mercieca P (2020) Social media and entrepreneurship research: a literature review. Int J Inf Manage 50(November 2018):90–110
42. Sivarajah U, Irani Z, Gupta S, Mahroof K (2020) Role of big data and social media analytics for business to business sustainability: a participatory web context. Ind Mark Manag 86(July 2018):163–179
43. Alonso D, Iborra A, Sanchez P, Requena F (2017) European entrepreneurial learning in information and communication technologies: the EU-xCEL experience in Spain. IEEE Glob Eng Educ Conf EDUCON 655–659
44. Rippa P, Secundo G (2019) Digital academic entrepreneurship: the potential of digital technologies on academic entrepreneurship. Technol Forecast Soc Change 146(July 2018):900–911
45. Bosma N, Hessels J, Schutjens V, Van Praag M, Verheul I (2012) Entrepreneurship and role models. J Econ Psychol 33(2):410–424
46. Zozimo R, Jack S, Hamilton E (2017) Entrepreneurial learning from observing role models. Entrep Reg Dev 29(9–10):889–911
47. Ferreto E, Lafuente E, Leiva JC (2018) Can entrepreneurial role models alleviate the fear of entrepreneurial failure? Int J Bus Environ 10(2):160–173
48. Androutsos A, Brinia V (2019) Developing and piloting a pedagogy for teaching innovation, collaboration, and co-creation in secondary education based on design thinking, digital transformation, and entrepreneurship. Educ Sci 9(2):1–11
49. Politis D (2005) The process of entrepreneurial learning: a conceptual framework. Entrep theory Pract 29(4):399–424
50. Browder RE, Aldrich HE, Bradley SW (2019) The emergence of the maker movement: implications for entrepreneurship research. J Bus Ventur 34(3):459–476
51. Malhotra A, Majchrzak A (2019) Greater associative knowledge variety in crowdsourcing platforms leads to generation of novel solutions by crowds. J Knowl Manag 23(8):1628–1651
52. Ben Arfi W, Hikkerova L (2019) Corporate entrepreneurship, product innovation, and knowledge conversion: the role of digital platforms. Small Bus Econ
53. Li L, Su F, Zhang W, Mao JY (2018) Digital transformation by SME entrepreneurs: a capability perspective. Inf Syst J 28(6):1129–1157
54. Yin Z, Gong X, Guo P, Wu T (2019) What drives entrepreneurship in digital economy? Evidence from China. Econ Model 82(September):66–73
55. Nambisan S (2017) Digital entrepreneurship: toward a digital technology perspective of entrepreneurship. Entrep Theory Pract 41(6):1029–1055

The Soft Side of Technology Adoption: A Study of Customer Experience on Mobile Phone Usage

Yet Mee Lim, Tat-Huei Cham, Jason Wai Chow Lee, and Chuen Kee Pek

Abstract The smartphone industry has become a lucrative market attracting new entrants and resulting in an intense competition in the industry. In order for smartphones companies to sustain within the industry, it is essential for them to understand how consumers feel and experience with their products. It is critical to examine customer experience as it can influence consumers' perceived value of a product. The aim of this study is to investigate the relationships between five dimensions of customer experience and perceived value and brand loyalty in the mobile phone context. The results show that the sense, feel, act, and relate experience are significantly related to perceived value, which in turn, positively related to brand loyalty. Implications of the research findings are discussed.

Keywords Mobile phones · Smartphones · Customer experience · Technology adoption · Perceived value · Brand loyalty · Marketing

1 Introduction

Technology advancement is rapidly changing the way people utilize and experience technology. In fact, adoption of technology has become a necessity in everyday life. Due to such technology advancement, the usage of mobile phones has become increasingly important among the consumers. Generally, smartphones are mobile phones with more advanced computing ability and functions which comprise of media player, camera, messaging, recorder, calculator, calendar, and notepad.

Y. M. Lim · T.-H. Cham (✉) · J. W. C. Lee · C. K. Pek
UCSI Graduate Business School, UCSI University, Kuala Lumpur, Malaysia
e-mail: LimYM@ucsiuniversity.edu.my

J. W. C. Lee
e-mail: wclee@ucsiuniversity.edu.my

C. K. Pek
e-mail: pekck@ucsiuniversity.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_67

Smura et al. [1] concluded that people are becoming inseparable from their smartphones as they carry them all the time.

In the smartphone industry, Samsung is the industry leader of Android operating system as it is the largest contributor of market share in the global operating system [2]. Unlike Apple, Samsung's strategy is to target the market with broad ranges and varieties of smartphone, ranging from low-end phones to high-end phones. Its low-end and mid-range smartphones are affordable to those not-so-rich people. In contrast, its high-end smartphones, targeting mainly on more affluent buyers, offer attractive features with more benefits and higher quality at a relatively higher price which. Samsung has also been implementing market expansion strategies with the goal to boost its market share in the smartphone industry [3].

Due to shorter life cycle of mobile telecommunication products and higher customers' product expectations, customer loyalty has become the foundation of company to obtain sustainable competitive advantage for survival [4]. Hence, it is crucial for the mobile phone companies to explore the impact of customer experience on perceived value and product brand loyalty, so that they can remain competitive by tailoring appropriate experiential marketing strategies.

There are many studies conducted to investigate the adoption of technologies from the system (hardware) perspective in relation to its ease of use and perceived usefulness [5, 6]. However, the application of customer experience in the adoption of technology (the soft side) is relatively new. This knowledge gap represents a research deficiency that require further investigation. Hence, this study is one of a few that aims to examine customer experience in relation to the use of Samsung smartphones, given its popularity as mentioned above. Specifically, the objective of this study is to investigate which of the customer experience dimensions (feel experience, act experience, sense experience, relate experience, and think experience) would influence perceived value and brand loyalty among Samsung smartphones users. The research findings of this study will offer greater insights on the smartphone usage experience, which are valuable inputs for smartphones manufacturers.

2 Literature Review

A well-known concept of experiential marketing is the strategic experiential modules by Schmitt [7] and [8]. According to [7], strategic experiential modules can be used by marketers to develop various types of customer experiences. This module includes five dimensions of experiences, which are feel experience, act experience, sense experience, relate experience, and think experience [9]. Human's five senses (sight, hearing, taste, scent, and touch) will contribute to sense experience. These five senses can influence customers' desire and intention to purchase as well as affect their brand loyalties. Feel experience consists of two basic elements, which are mood and emotions. Mood is human's state with unidentifiable reason whereas emotion is human's state with identifiable factors ranging from joy, anger, envy, pride, and so

on. The emotion and moods during the consumption of products can influence the feel experience of the consumers. Think experience can be defined as an experience to stimulate consumers' creative and innovative thinking about company's products. It can create perceptions and solve consumers' problems in a creative method. Think experience can be stimulated when a new idea, feature, or product is being introduced. Act experience refers to the social interaction experience that deals with consumers' lifestyle, physical body, and behaviour. Throughout the act experience, a positive relationship between users and the products can be developed. Hence, act experience is an experience concerning the lifestyle and social behaviour of the consumers. Relate experience is an experience which consists of consumers' connections with others through the purchase and consumption of certain brands. Relate experiences enables consumers to build social connection with the communities to develop brand preferences for a certain product.

2.1 Sense Experience and Perceived Value

Various studies have indicated the relationship between sense experience and perceived value. As stated by Ziethaml [10], perceived value is consumers' overall evaluation of products utilisation, based on what is received (benefits) and what is given (costs and sacrifices). Perceived value is the main criteria for a customer to purchase a product. Customers expect a value in return for the amount of money they spent [11–13]. Lin and Chiang [14] found that aesthetic design or ambient cues of store environment mediated by positive emotions will influence consumers' perception on price, making them pleased to pay more for the products. Alagoz and Ekici [15] also depicted that senses such as music, lightning, and heat in the plane positively related to experiential brand perception of vacation; and passengers would likely to pay more for the flight tickets. Firms often display products and decorate their stores to satisfy visitors' sense of sight, giving off a perception of elegance to visitors [16, 17]. Moreover, [18] stated that physical features would shape tourists' image towards the night market as they will evaluate and form overall value perceptions after strolling in the night market. For written materials, attractive contents in blogs can trigger positive sense experience [19]. Thus, we hypothesized that:

H1: There is a positive relationship between sense experience and perceived value.

2.2 Feel Experience and Perceived Value

Most feel experiences are generated during the use of a product or being presence in a service environment. Researchers have found that feelings of pleasure induced by favourable music will increase consumers' time spent in the store and reduce their

perceived price and time cost. Positive emotions like happiness and excitement were found to create memorable consumption experience, and such pleasurable experience is associated with customer perceived value and satisfaction [14, 15, 20, 21]. Consumer will interpret the environment and product cues both cognitively and emotionally, and the feelings evoked would have a positive impact on the perceived value of the product and brand engagement [22–24]. Hence, the following hypothesis was stipulated:

H2: There is a positive relationship between feel experience and perceived value.

2.3 Think Experience and Perceived Value

Think experience is the way of how customers think about the product. Marketers tend to use a creativity way in encouraging customers to think about their products by increasing consumers' understanding and knowledge about the products, which in turn enhance customer's perceived value [25, 26]. A differentiation strategy is to encourage people to think about the product features that differentiate its brand from those of other competitors, with the aim to increase perceived value [27]. Perceived value can be enhanced by influencing customer's thinking pattern, attracting their interest towards the product by increasing their curiosity for the product [28]. The curiosity creates a desire to investigate the product capabilities which could meet customers' expectation, thus leveraging perceived value of the product throughout the ongoing investigation [29, 30]. This leads to the formulation of the following hypothesis:

H3: There is a positive relationship between think experience and perceived value.

2.4 Act Experience and Perceived Value

According to [31], act experience engages consumers in physical activities and changes their lifestyles and behaviours. It improves consumers' personal life quality which in turn, induces a favourable value perception on the company's products and services. Alagoz and Ekici [15] found a positive relationship between physical experience and vacation perception. Passengers' vacation purpose will change their value perception; for instance, business travellers would have a high perceived value on flight services offering work-enhancement facilities like communication tools. Chen and Lin [29] stressed that mobile services and mobile marketing enhance consumers' daily experiences resulted from the usage of phones, and daily experiential consumption of mobile phones creates value. The act experience of phones integrates several behavioural approaches, including physical activities and living patterns [28]. Behavioural consumption practices such as owning and using

phones to coordinate everyday activities would create perceived value of the phones. Thus, the following hypothesis was formed:

H4: There is a significant relationship between act experience and perceived value.

2.5 Relate Experience and Perceived Value

Relate experience appeals to consumers' needs for interacting and belonging to a subculture or community. Ownership of a product provides a sense of individual self-realisation or social identity by relating own self to other social groups or cultures [25]. As global young consumers utilise the same social media platforms, the experiences from using the same ICT device and social media platforms might account for universal phenomena of belongingness and self-esteem [32]. This indicates that relate experience creates values linking to interpersonal relations, social identity, and self-realization. It would also positively impact consumers' perceived return on investment, wherein value or benefits gained is relatively larger than the cost of investment, equivalent to a higher perceived value [33]. Hence, it was hypothesized that:

H5: There is a positive relationship between relate experience and perceived value.

2.6 Perceived Value and Brand Loyalty

Brand loyalty is the deep commitment to re-buy or re-patronise a preferred product or service consistently in the future. It refers to repurchasing a preferred product in a consistent manner, regardless of the environment influences or any marketing efforts [34]. According to [35], higher perceived value on a brand reinforces consumers' trust in that particular brand. Such high brand trust would lead to a strong brand loyalty. Various researchers have established the causal link among brand awareness, association, perceived value, and brand loyalty [36–38]. Thus, we hypothesise that:

H6: There is a positive relationship between perceived value and brand loyalty.

The following depicts the research model of this study (Fig. 1):

3 Research Method

The target respondents of this study were university students and users of Samsung smartphones. To expand social networks and stay connected with friends, the usage of technological products, especially smartphones, has thus increased dramatically

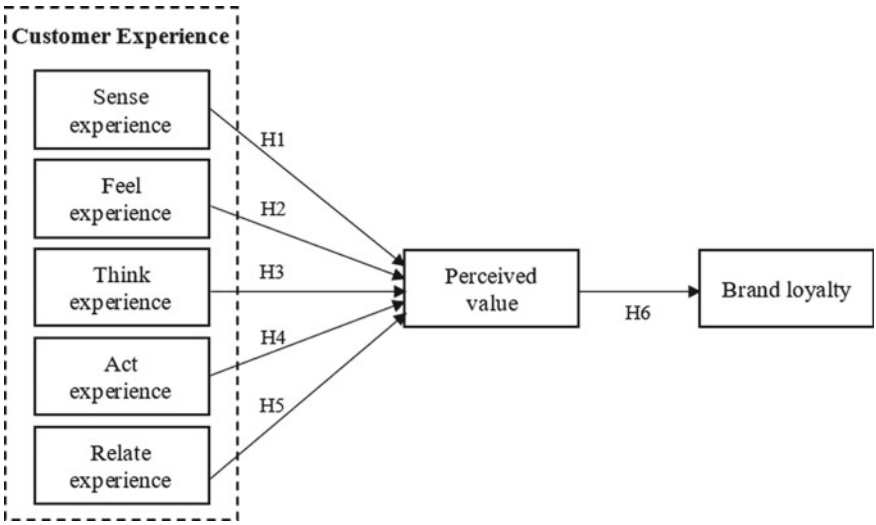


Fig. 1 Research model

among the university students. University students, who are mostly fall in the category of Generation Y, are the most important target markets in the smartphone industry, as they view smartphones as a necessary item in their lives [39–41]. Samsung has been chosen for this study as it holds the biggest market share in the smartphone industry and Android operating system market [2].

Questionnaires created in the Google Form platform were distributed to the students of a private university via Facebook, a popular social media in Malaysia. The questionnaire consisted of two sections. Section A included screening questions to filter out responses of non-targeted respondents. Students who were non-Samsung users would not be included in the survey. This section also included some demographic questions regarding respondents' age, year of study, major of study, and some general questions related to smartphone ownership period and previous used smartphone brands.

Section B included the measurement items of the five dimensions of customer experience, perceived value, and brand loyalty. All measurement scales were adapted from past studies. The measurement items of customer experience dimensions were adapted from the studies of [42–46]. Items measuring perceived value were adapted from [47], and the items of brand loyalty were adapted from the study of [11] and [48].

A total of 291 responses were received from the target respondents. Out of these 291 responses, 26 were excluded due to extremely biased responses yielding a sample size of 265. questionnaires were usable. A sample size of 265 is considered large sufficient for a multivariate research study [49–51]. Moreover, the issue of common method variance (CMV) for this study was addressed based on the guideline recommended by Memon et al. [52]. According to [52], CMV can be

addressed using Harman's single factor test that is based upon factor analysis. The output from the factor analysis indicated that the first and largest variance explained was found to be less than 40%. Moreover, the model with all the indicators modelled to one factor was found not to be fit. Hence, there was no indication of common method bias.

The present study uses the confirmatory factor analysis (CFA) to assessed the validity for the variables used. As suggested by Kim and Cho [49], the primary step of CFA is to determine the model fit values for the measurement model before proceed with the validity analyses (i.e. discriminant and convergent). Kim and Cho [49] suggested that a proposed research model is considered fit if the values of GFI is greater than 0.90, TLI is greater than 0.90, PNFI is greater than 0.50, Normed Chi-square (χ^2/df) is less than 3, and RMSEA is below 0.08. The assessment of the measurement model in this study indicated that the GFI = 0.820, TLI = 0.923, PNFI = 0.779, χ^2/df = 1.992, and RMSEA = 0.0461, which proposed that the model is fairly fit.

The convergent validity in this study was evaluated based upon the conditions suggested by Kim and Cho [49]. The criteria to determined convergent validity includes the value for all the indicators of the constructs must above 0.60, and the value of average variance extracted (AVE) and composite reliability for the constructs should be larger than the cut-off value of 0.50 and 0.70 respectively. The outcome of the analysis indicated that two items from act experience had been dropped from further analysis due to its low loadings (e.g. <0.60). Subsequently, convergent validity analysis was conducted after the removal of the items with low loadings. The outcome in the Table 1 shows that the measurement model satisfied all the criteria suggested by Kim and Cho [49], which indicated that the convergent validity was confirmed for the present study.

In addition, the discriminant validity for this research was addressed in referring to the recommendation by MacKenzie and Podsakoff [53]. According to [53], discriminant validity for the variables can be assessed upon the comparison of (1) AVE against the value of maximum shared variance (MSV) and (2) squared root of AVE against the variance between variables. It was reported that discriminant validity is considered to be confirmed if the value of AVE of a particular variable is greater than its respective MSV value, and the value of squared root of AVE for the

Table 1 Result of convergent validity

	Items	Loadings	AVE	CR
Perceived value	5	0.757–0.826	0.631	0.895
Sense experience	4	0.729–0.821	0.641	0.877
Feel experience	6	0.653–0.813	0.545	0.877
Think experience	4	0.654–0.826	0.534	0.820
Act experience	6	0.672–0.877	0.651	0.881
Relate experience	6	0.756–0.947	0.762	0.950
Brand loyalty	5	0.862–0.934	0.811	0.955

Notes: *CR* Compostite relaibility; *AVE* Average variance extracted

Table 2 Result of discriminant validity

	MSV	AVE	1	2	3	4	5	6	7
Perceived value	0.540	0.631	*0.795*[b]						
Sense experience	0.520	0.641	**0.595**[a]	*0.801*					
Feel experience	0.520	0.545	**0.651**	**0.721**	*0.738*				
Think experience	0.320	0.534	**0.457**	**0.538**	**0.566**	*0.731*			
Act experience	0.440	0.651	**0.663**	**0.578**	**0.637**	**0.469**	*0.807*		
Relate experience	0.383	0.762	**0.548**	**0.421**	**0.483**	**0.317**	**0.619**	*0.873*	
Brand loyalty	0.540	0.811	**0.735**	**0.539**	**0.573**	**0.367**	**0.463**	**0.437**	*0.900*

Notes: *MSV* Maximum shared variance; *AVE* Average variance extracted
[a]The off-diagonal values indicates the variance shared between variables
[b]The diagonal values indicates the values of squared root average variance extracted by the variable

individual variables are larger than the variance shared between them. The outcome as illustrated in Table 2 found that the value of AVE for all the variables are greater than the values of MSV of its owns and the values of squared root of AVE are greater than the values of correlation between any variables, which suggesting that discriminant validity for the present study was established.

As for the structural model, the causal paths in the research model were addressed using path analysis. The outcome of the analysis on the structural model denoted that the GFI = 0.818, TLI = 0.923, PNFI = 0.785, χ^2/df = 1.997, and RMSEA = 0.061, which indicating that the structural model in this case is relatively fit. Moreover, the outcome of the path analysis indicated that all the hypotheses except H3 were found to be significant and supported. The finding in Table 3 shows that sense experience ($\beta = 0.169$, $p < 0.05$), feel experience ($\beta = 0.112$, $p < 0.05$), act experience ($\beta = 0.266$, $p < 0.001$), and relate experience ($\beta = 0.176$, $p < 0.05$) were significantly related to customers' perceived value. Moreover, the path analysis also indicated that perceived value was stated to have a significant positive influence on brand loyalty ($\beta = 0.744$, $p < 0.001$).

Table 3 Result of hypotheses testing

	Standardised estimate (β)	Critical ratio	Hypothesis
H1: Sense experience → Perceived value	0.169	0.082*	**Yes**
H2: Feel experience → Perceived value	0.266	0.112*	**Yes**
H3: Think experience → Perceived value	0.040	0.066 [ns]	**No**
H4: Act experience Perceived value	0.266	0.073**	**Yes**
H5: Relate experience → Perceived value	0.176	0.049*	**Yes**
H6: Perceived value → Brand loyalty	0.744	0.095**	**Yes**

Notes: **p-value < 0.001, * p-value < 0.05, ns = not signficant

4 Discussion and Conclusion

The results of this study indicate that sense experience, feel experience, act experience, and relate experience will positively influence university students' value perception on Samsung smartphones. Perceived value was also reported to have a significant positive impact on brand loyalty in this case. In other words, to create brand loyalty, it is necessary for Samsung to produce smartphones that could build consumers' perceived value on the products by stimulating users' sense, feel, act, and relate experience favourably.

First of all, Samsung producers must be creative in designing their products with models of chic and modern design. The ability to provide a modern and advance design that appeals to all of the five sense of the users will definitely create a competitive advantage. Such an appealing phone product will allow users to relate well to other people through ownership. The smartphones should be able to enhance users' relationship with others by making users engage in social activities and networking more often. Most importantly, marketers should ascertain that their smartphones arouse users a sense of belonging or esteem by gaining social recognition and status upon owning a high-end Samsung smartphone.

By aligning the superior smartphone functions and specifications with the target consumers' needs and behavioural lifestyle, Samsung marketers could trigger feel experience among consumers as they enjoy and are excited by the interesting functions. Since users seek to feel comfortable and pleasure in using smartphone with good mood, Samsung marketers should highlight the smartphones' ease of use, including simplicity of usage, easy-to-learn usage steps, convenience, and time-saving. The advertising of Samsung smartphones should also display positive emotions like joyfulness and happiness, conveying the message to the audience that they would feel the same while using Samsung smartphones.

Furthermore, the use of smartphones should enable users to engage in many daily activities as possible to enhance act experience. To target the university students, Samsung marketers should thus research on their usages of smartphones. They should emphasise the ability of its smartphones to solve the users' problems. The hardware of Samsung smartphones must be compatible with various productivity applications; and the smartphones can be designed as an alternative tablet or laptop.

Future research may examine the user experience of other smartphones of different makes for comparison purpose. The differences in expected experience among students and non-students and among users of different age groups are worth examining to tailor specific designs for specific groups. Despite the limitations, this study has provided some insights to the smartphone industry leaders on designing and promoting their products to maximize the experience of the users for ultimate satisfaction.

References

1. Smura T, Kivi A, Toyli J (2009) A framework for analyzing the usage of mobile services. Info J Policy, Regul Strategy Telecommun 11(4):53–67
2. Tech.co Homepage, https://tech.co/phones/best-phone-brands. Last Accessed on 2021/01/02
3. Businessinsider.com Homepage, https://www.businessinsider.com/samsung-corporate-strategy-2013-3. Last Accessed on 2020/09/03
4. Haverila MJ, Haverila KC (2015) Brand satisfaction and repurchase intent in the cell phone product market. Acad Mark Stud J 19(1):197–212
5. Cheng CH, Chen CH, Chen YS, Guo HL, Lin CK (2019) Exploring Taiwanese's smartphone user intention: an integrated model of technology acceptance model and information system successful model. Int J Soc Humanistic Comput 3(2):97–107
6. Rigopoulou ID, Chaniotakis IE, Kehagias JD (2017) An extended technology acceptance model for predicting smartphone adoption among young consumers in Greece. Int J Mobile Commun 15(4):372–387
7. Schmitt BH (1999) Experiential marketing. J Mark Manage 15(13):53–67
8. Schmitt BH (1999) Experiential marketing: how to get customers to sense, feel, think, act and relate to your company and brand. New York
9. Hultén B (2017) Branding by the five senses: a sensory branding framework. J Brand Strategy 6(3):281–292
10. Ziethaml VA (1988) Consumer perceptions of price, quality, and value: a means- end model and synthesis of evidence. J Mark 52(3):2–22
11. Machado R, Cant MC, Seaborne H (2014) Experiential marketing on brand advocacy: a mixed-method approach on global apple product users. Int Bus Econ Res J (IBER) 13(5):955–962
12. Cham TH, Lim YM, Sia BC, Cheah JH, Ting H (2020) Medical tourism destination image and its relationship with the intention to revisit: a study of Chinese medical tourists in Malaysia. J China Tourism Res. https://doi.org/10.1080/19388160.2020.1734514
13. Cham TH, Cheng BL, Low MP, Cheok JBC (2020) Brand image as the competitive edge for hospitals in medical tourism. Eur Bus Rev. https://doi.org/10.1108/EBR-10-2019-0269
14. Lin MQ, Chiang YF (2010) The influence of store environment on perceived experiential value and behavior intention. Asia Pac Manag Rev 15(2):281–299
15. Alagöz SB, Ekici N (2014) Experiential marketing and vacation experience: the sample of Turkish airlines. Procedia Soc Behav Sci 150:500–510
16. Fujiwara K, Nagasawa S (2015) Relationship between purchase intentions for luxury brands and customer experience: comparative verification among product categories and brand ranks. Sci J Bus Manage 3(2–1):1–10
17. Cham TH, Easvaralingam Y (2012) Service quality, image and loyalty towards Malaysian hotels. Int J Serv Econ Manage 4(4):267–281
18. Lee SH, Chang SC, Hou JS, Lin CH (2008) Night market experience and image of temporary residents and foreign visitors. Int J Cult, Tourism Hospitality Res 2(3):217–233
19. Hsu HY, Tsou HT (2011) Understanding customer experiences in online blog environments. Int J Inf Manage 31(6):510–523
20. Bagdare S, Jain R (2013) Measuring retail customer experience. Int J Retail Distrib Manage 41(10):790–804
21. Lim XJ, Ng SI, Chuah F, Cham TH, Rozali A (2019) I see, and I hunt: The link between gastronomy online reviews, involvement and behavioural intention towards ethnic food. British Food J 122(6):1777–1800
22. Rozita NM, Halimahton B (2014) Exploring customer attachment behaviour for sustaining retail industry in Malaysia. World J Entrepreneurship, Manage Sustain Dev 10(1):69–76

23. Rozita NM, Rosidah M, Ramesh K, Shafinar I (2015) The modern retail customer's experience on customer engagement: evidence from health and personal care stores in Malaysia using structural equation modelling approach (SEM). Global J Bus Soc Sci Prev 1 (2):441–448
24. Sheng ML, Teo TS (2012) Product attributes and brand equity in the mobile domain: the mediating role of customer experience. Int J Inf Manage 32(2):139–146
25. Nagasawa S (2008) Customer experience management: influencing on human Kansei to management of technology. TQM J 20(4):312–323
26. Cham TH, Cheng BL, Ng CKY (2020) Cruising down millennials' fashion runway: a cross-functional study beyond Pacific borders. Young Consumers. https://doi.org/10.1108/YC-05-2020-1140
27. Head M, Ziolkowski N (2012) Understanding student attitudes of mobile phone features: rethinking adoption through conjoint, cluster and SEM analyses. Comput Hum Behav 28 (6):2331–2339
28. Chen SC, Lin CP (2015) The impact of customer experience and perceived value on sustainable social relationship in blogs: an empirical study. Technol Forecast Soc Chang 96:40–50
29. Andrews L, Drennan J, Russell-Bennett R (2012) Linking perceived value of mobile marketing with the experiential consumption of mobile phones. Eur J Mark 46(3/4):357–386
30. Liu F, Zhao X, Chau PY, Tang Q (2015) Roles of perceived value and individual differences in the acceptance of mobile coupon applications. Internet Res 25(3):471–495
31. Liu Y, Stoel L (2013) Winning consumers through experience: competition between domestic and international businesses. Am J Ind Bus Manag 3:341–348
32. Kang S, Jung J (2014) Mobile communication for human needs: a comparison of smartphone use between the US and Korea. Comput Hum Behav 35:376–387
33. Chen YK, Hsieh T (2010) A study of the relationship among experiential marketing, experiential value and customer satisfaction. J Stat Manag Syst 13(6):1283–1303
34. Cheng BL, Cham TH, Micheal D, Lee TH (2019) Service innovation: building a sustainable competitive advantage in higher education. Int J Serv, Econ Manage 10(4):289–309
35. Mishra A, Dash SB, Cyr D (2014) Linking user experience and consumer—based brand equity: the moderating role of consumer expertise and lifestyle. J Prod Brand Manage 23(4/5):333–348
36. Chang HH, Wang HW (2011) The moderating effect of customer perceived value on online shopping behaviour. Online Inf Rev 35(3):333–359
37. Pappu R, Quester PG (2008) Does brand equity vary between department stores and clothing stores? results of an empirical investigation. J Prod Brand Manage 17(7):425–435
38. Lin J, Lobo A, Leckie C (2017) The role of benefits and transparency in shaping consumers' green perceived value, self-brand connection and brand loyalty. J Retail Consum Serv 35:133–214
39. Mohd Suki N (2013) Students' demand for smartphones: Structural relationships of product features, brand name, product price and social influence. Campus-Wide Inf Syst 30(4):236–248
40. Ting DH, Lim SF, Patanmacia TS, Low CG, Ker GC (2011) Dependency on smartphone and the impact on purchase behaviour. Young Consumers: Insight Ideas Responsible Marketers 12(3):193–203
41. Cheah JH, Ting H, Cham TH, Memon MA (2019) The effect of selfie promotion and celebrity endorsed advertisement on decision-making processes: a model comparison. Internet Res 29 (3):552–577
42. Maghnati F, Ling KC, Nasermoadeli A (2012) Exploring the relationship between experiential marketing and experiential value in the smartphone industry. Int Bus Res 5 (11):169–177
43. Khan I, Rahman Z (2014) Influence of experiential marketing on customer purchase intention: a study of passenger car market. Manage Labour Stud 39(3):319–328

44. Kazutoshi F, Shin'ya N (2015) Relationship between purchase intentions for luxury brands and customer experience: Comparative verification among product categories and brand ranks. Sci J Bus Manage 3(2–1):1–10
45. Cleff T, Lin IC, Walter N (2014) Can you feel it? The effect of brand experience on brand equity. IUP J Brand Manage 11(2):7–27
46. Yang ZY, He LY (2011) Goal, customer experience and purchase intention in a retail context in China: an empirical study. Afr J Bus Manage 5(16):6738–6746
47. Dodds WB, Monroe KB, Grewal D (1991) Effects of price, brand, and store information on buyers' product evaluations. J Mark Res 28:307–319
48. Kim A, Cho YC (2015) Exploring antecedents and consequences of customer satisfaction with smartphones: Implications for CRM. J Bus Econ Res 13(4):215
49. Hair JFJ, Black WC, Babin BJ, Anderson RE, Tatham RL (2010) Multivariate data analysis a global perspective. Pearson Education International, New Jersey
50. Saunders M, Lewis P, Thornhill A (2012) Research methods for business students. 6th edn. Pearson, UK
51. Memon MA, Ting H, Cheah JH, Ramayah T, Chuah F, Cham TH (2020) Sample size for survey research: review and recommendations. J Appl Struct Equ Model 4(2):1–20
52. MacKenzie SB, Podsakoff PM (2012) Common method bias in marketing: causes, mechanisms, and procedural remedies. J Retail 88(4):542–555
53. Fornell C, Larcker DF (1981) Structural equation models with unobservable variables and measurement error: algebra and statistics. J Mark Res 18(3):382–388

Artificial Intelligence Applications in Halal Tourism to Assist Muslim Tourist Journey

Mohamed Battour, Khalid Mady, Mohamed Elsotouhy, Mohamed Salaheldeen, Israa Elbendary, Mohamed Marie, and Idris Elhabony

Abstract Using automated services based on artificial intelligence (AI) has recently witnessed a tremendous leap across various industries. While the implementation of intelligent automation in the tourism and hospitality industry is increasing nowadays, Halal Tourism still needs more investigations. Therefore, the aim of this study is to explore how AI can be used to ameliorate the Muslim tourism experience during the tourism journey. This conceptual paper reviewed AI's role in every phase of the tourism journey: planning phase, staying phase, and evaluation phase which could be utilized to improve Muslim tourism experience. The implications of this study appear theoretically and practically. This conceptual paper tried to fill the gap relevant to AI research in Halal tourism, which is very limited, besides providing an overview of how AI can be used professionally in the tourism and hospitality industry.

M. Battour (✉)
College of Business Administration, University of Sharjah, Sharjah, UAE
e-mail: mbattour@sharjah.ac.ae

K. Mady
Faculty of Business and Economics and social development, Universiti Malaysia Terengganu, Kuala Terengganu, Malaysia
e-mail: Khaled.Mady@com.kfs.edu.eg

M. Elsotouhy
Faculty of Commerce, Mansoura University, Mansoura, Egypt

M. Salaheldeen
Faculty of Economics and Muamalat, Universiti Sains Islam Malaysia (USIM), Nilai, Malaysia

I. Elbendary · M. Marie
Faculty of Commerce, Cairo University, Giza, Egypt
e-mail: Israa_mohamed@foc.cu.edu.eg

M. Marie
e-mail: Mohamed_marei@foc.cu.edu.eg

I. Elhabony
Idris Abdelgawad Elhabony, Faculty of Economic, University of Benghazi, Benghazi, Libya

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_68

Keywords Halal tourism · Artificial intelligence · Digital platforms

1 Introduction

The hospitality and tourism industry has recently been popular with the usage of smart services that are featured with innovation in technology and observed escalation in the application of artificial intelligence. Artificial intelligence (AI) refers to the emulation of human intelligence in computers designed to think and imitate their behavior like humans. Indeed, one of the critical factors of digital transformation witnessed in many sectors is automation. Its main idea is not limited to replacing the human effort with machines but rather integrating machines into a self-governance system that can carry out an entire process without a human's assistance [1]. At the same time, tourists demand more personalization in the offered services and seek a better digital experience. The challenge is to combine their real-time trip with a tailor-made experience [2].

The destination marketers that target Muslim tourists as an important segment in the global travel market should cater to this segment's specific needs [3–5]. Therefore, in the pre-trip stage, Muslim tourists can have a unique and inspiring travel experience if intelligent automation and AI are applied. Further, these technologies can facilitate getting needed information, booking, and pre-arrival experience. In addition, in the pre-trip stage, AI makes it easier for destination marketers to increase the automated tourism experience for Muslim tourists during the staying stage to get feedback after leaving the destination for improving that experience. Besides, AI can facilitate understanding Muslim tourists' specific needs and the quick response to Halal tourism businesses' challenges. Indeed, it can bring remarkable, tailor-made experience and offer the chance to excel in the exploration of a Halal tourism destination. The tools of AI can warrant the efficiency, innovation, and sustainability of Halal tourism providers [6].

The areas of AI, robotics, and automation have been the concern of tourism research; however, contribution in this area is limited to describing what is currently applied and what are the recommendations for the future [7–11]. It is noted that there is a lack of theoretical and empirical research on the application of intelligent automation in the hospitality and tourism industry, specifically in Halal tourism. According to Tussyadiah [12], studying automated tourism experience is needed. AI and its applications have become crucial for providing the tourism and hospitality industry with effective solutions to improve tourists' experience [13]. Therefore, this conceptual paper aims to review AI's role in every phase of the tourism journey: planning phase, staying phase, and evaluation phase which could be utilized to improve Muslim tourist experience. Critical assessment in Halal tourism and AI literature is used to explore AI applications in Halal Tourism to assist Muslim Tourist Journey. This paper is also expected to provide some suggestions for tourism operators and service providers to be able to cater for Muslim tourists using AI.

2 Artificial Intelligence in the Tourism and Hospitality Industry

Establishing unique tourism experiences requires the involvement of many factors, which are developed according to the deep understanding of tourists' demands, to make sure that tourism experience is not only seamless but more likely memorable [14]. Emerging new technologies powered by AI, such as virtual guide and virtual host, Chatbots, smart room, and digital personal assistant, are drastically transforming the tourism experience [12]. These technologies have become megatrends that make the travel experience more affordable, well-organized, and accessible to many people [15]. Intelligent automation has brought substantial benefits to both tourists and destinations [12]. The synergy between human efforts and intelligent systems can increase productivity, increase flexibility in the workplace, and enhance decision-making efficiency. Additionally, service providers are trying to exploit these technologies to design and deliver a unique customer experience that leads to promising business outcomes and provides them with a competitive advantage.

In the hospitality industry, data mining and intuitive computing are used by AI agents to assist tourists in choosing the right hotel. Moreover, when the tourist arrives at the hotel room, a digital camera recognizes his face using facial recognition technology, then the door opens automatically. In the room, the temperature is determined based on tourist preferences, which are sent automatically with the location and estimated arrival time to the hotel through the mobile phone. The tourist is welcomed by a robot that assists him/her in unpacking the bag and ironing rumpled clothes. Hotel appliances such as TV, lights, and curtains are operated with merely a voice command [16].

All these features may seem far-fetched and modern for some, whereas many hotels adopted this technology for better and smart hotel services [16]. For instance, in Peninsula Chicago, tourists can personalize the service they receive using a "smart-mirror" with the features of a tablet in Starwood Hotels. In Marriott, too, guests personalize their itinerary using an app. Likewise, Hilton is planning to introduce a "smart-room" that is connected to a computer system. In addition, other industries began to apply technology and smart services as well. Such destinations, Singapore and Malta, created smart travel destinations using smart facilities and services. Similarly, many airports worldwide use connected Bluetooth devices to invigilate real-time traveler's flow and attitude to minimize crowds and offer personalized notifications and smart services [16].

It is expected that the application of AI and robotics in tourism marketing will improve and enhance travelers' overall experience. Furthermore, chatbots, virtual reality, and language translators can be employed in the Travel, Tourism and Hospitality industry [17]. It is indisputable that technological progress is unstoppable. Robots, AI algorithms, chatbots, and kiosks will improve technical characteristics and will be more affordable, which will make them more widely used in the tourism and hospitality industry [18, 19]. Travel, tourism, or hospitality

(TTH) company robots, self-service kiosks, and AI software packages are considered assets [20]. Therefore, a company would invest in them if the financial performance improves and creates value for their shareholders [21].

AI opens up new and exciting opportunities for value co-creation by knowing customer needs and preferences and consequently enabling resource integration between service providers and customers [22, 23]. However, the infusion of AI into service renders value creation a more complicated process whereby coordinated, harmonized resource inputs assume even greater importance [24]. In addition, AI-powered interactions can also fail, potentially leading to anger, confusion, and customer dissatisfaction. The application of intelligent automation might pose a degree of risk that can threaten the tourist experience and future tourism industry in the smart service era, including social interaction challenge [25], security and privacy challenges [12], and socio-ethical challenge such as the fear of job loss acceptance of AI technologies in the workplace [25].

3 Artificial Intelligence in Halal Tourism

Tourism recognizes the rising interest in Halal tourism from both an industrial and research point of view. In tourism literature, the 'Halal tourism' terminology was first used by Battour [26, 27]. Recently, Halal tourism is defined by Battour et al. [32] "as any tourism object or action that is permissible according to Islamic teachings to satisfy the Muslim traveler's needs and achieve the destination's Muslim-friendliness". Before using 'Halal Tourism' Term in the first paper published in Scopus journal in 2010 [27], the standard terminology was 'Islamic Tourism,' used by Henderson [28]. Halal tourism concept goes around Muslim travels' needs as sometimes are different compared to other tourists needs. Due to the absence of these needs, Muslim tourists may decide not to travel to a specific destination. For example, the absence of Halal food and Halal restaurant could be a source of worry to Muslim travelers and to those with whom they interact. Moreover, some other Muslim traveler's needs might not be easy to be provided by tourism operators and/or tourism service providers such as sharia-compliant hotels where there is a ban on alcoholic drinks, Muslim prayer facilities, separated swimming pools, and gymnasium for men and women, Islamic entertainments i.e. TV channel, and separated/family beaches.

The activities of Halal tourism may be seen as restrictions on the growth of tourism destinations. Such constraints are significant and can serve as major challenges to the planning and marketing of tourism in Muslim and non-Muslim destinations. Therefore, some other terminologies relating to Halal tourism in the Muslim travel market have been used, such as Halal/Muslim travel, Muslim friendly tourism, and Halal friendly destination [3–5, 29–35]. In non-Muslim countries including Japan, Australia, and New Zealand, some tourism activities related to these concepts are applied [31, 36–38]. In Halal tourism, Muslim-friendly terminology may be more fitting when non-Muslim destinations attempt to portray

themselves as Muslim friendliness [32]. In the definition of Halal tourism in previous literature, friendliness is ignored [For example, 3–5, 35, 39, 40–44]. The key challenge is to find a balance between the catering needs of Muslim and non-Muslim visitors without interfering with Islamic teachings. Therefore, AI and its applications could help tourism destinations that are willing to adopt Halal-friendly tourism to satisfy Muslim travel needs effectively and timely [13]. AI will play a vital role in developing and marketing the tourist's products and services. The future of tourism industry will consequently witness a quantum leap in the exploitation of intelligent technology applications such as Intelligent Personal Assistant (IPA), Remote Control Tourists (RCT), Human–computer interaction, and Augmented Reality (AR). The same applies to Halal tourism, where intelligent solutions and environments should be found to enhance Muslim tourists' experience at each point of their trip—from planning to share their travel feedback [45].

3.1 Planning Phase

The most popular medium for travelers to plan and prepare their journeys for either pure leisure or business purpose is the Internet and mobile applications [46]. In the online environment, specialty platforms, and online reviews on social networks can provide detailed information that helps potential travelers and tourists make decisions on destinations and leisure activities [47]. With the magnitude and variety of this information, travel planning has become a challenging task consuming much time [13, 47]. AI provides effective solutions to information overload, such as social media analytics and tourism recommender systems [48]. Tourists can also obtain instant information and assistant by AI-powered tools such as chatbots [7].

In order to provide a research and data analysis platform for Halal travel in hospitality and tourism industry, AI applications are needed. In this category, data analytics providers which use AI are scarce. CrecsentRating, which is headquartered in Singapore, is the only company offering this service. In order to compete in this market, Halal Travel needs more AI applications. By analyzing Muslim travelers' preferences and needs, the study and data analysis platform will play an important role. In addition, the Halal Booking Platform is also important. There are limited companies, such as Halalbooking and HalalTrip, that offer Halal booking services. Using AI applications in this segment will help Muslim tourists choose a relevant destination and travel packages in the planning phase advertisements on companies' websites and social media networks. Moreover, by using AI algorithms on the basis of behavioral characteristics, the target prices that Muslim travelers are willing to pay could be personalized for them [49].

Social media platforms (e.g. Facebook, Twitter, and Instagram) are full of previous tourists' photos, ratings, and reviews for different destinations and brands. AI, in particular machine learning, has a significant role in analyzing the massive volume of social media data and transform it into meaningful and actionable information [50]. Muslim-friendly tourism is a fast-growing sector that requires

understanding Muslim tourists' unique needs and expectations and providing products and services upholding their principles and religions [45]. Following the data available on social media networks, AI and social media analytics can rapidly provide companies with in-depth insights into Muslim tourists' mindset and their attitudes and behavior [51]. Companies can tap into this information to help Muslim tourists choose a relevant destination and travel packages in the planning phase by analyzing advertisements on companies' websites and social media networks. Advancements in mobile computing, web technologies, social media platforms, and wireless networking lead to accurate tourist recommendations and effective and respect personal preferences [52]. A recommender system is an obvious example of AI application that is developed to provide users with active and personalized recommendations without needing a search query [48].

Tourism recommender system is very beneficial for Halal travel of interest to Muslim tourists are different from the other tourists in terms of several aspects such as food, attractions, lodging, and destinations. As such, tourism recommender systems play a crucial role in offering Muslim tourists relevant choices related to their profiles. Moreover, Chatbots and AI technologies transform the tourism industry, including Halal Tourism [12, 53]. A chatbot is a software program, sometimes called a robot or bot, which has a conversation in natural language with customers [54]. A chatbot is a feasible form of AI that concentrates on the user [55]. It occurs through textual or vocal format via unique gadgets such as Google, Siri, Amazon Alexa, computers, and smartphones [53]. According to Hsu et al. [56], the chatbot has several advantages: saving time, building a social relationship with consumers, building confidence with users, and fostering customers' emotional bonds with firms. The enormous advantages of chatbots appear in its ability to scheduling appointments, setting a reminder, booking tickets, and showing traffic and weather updates [55].

Several studies have revealed the importance of chotbot during the tourism journey. According to Tussyadiah [12], some AI-based software like chatbots can guide tourists throughout the whole journey, including pre-trip, real-time trip, and post-trip. In pre-travel, it can be used for several activities. American online travel shopping company Expedia, for example, utilized Facebook's technology to generate a basic bot to assist travelers to book hotels and booking rooms [57]. In the same vein, Ukpabi et al. [58] pointed out that chatbots can be utilized to enriching the pre-travel experience by allowing users to book rooms and other facilities, such as spa, airport tickets, and dinner reservations. According to Pillai and Sivathanu [53], chatbots are mainly used to plan the trips, support customers, provide recommendations, and supply suggestions to customers regarding travel issues. In Halal Travel, chatbots can help Muslim travel to identify many aspects relevant to his/her trip planning such as Halal airlines, availability of Halal food, prayer times, and Halal entertainment booking.

3.2 Staying Phase

In the staying phase, information gathered previously during the pre-trip stage, tourists will confirm it when they arrive at the destination. Recently, destination organizers, service providers, and tourists use technology to connect more easily for a better tourism experience during their stay [59]. Moreover, tourism chatbots can provide travel guidance 24 × 7, which will be helpful for the staying phase [53]. During the staying phase, customer can use chatbots to manage restaurant reservations through exploring restaurant photos, menus, price, and available tables, as well as change or cancel the reservation [57]. Because of machine learning and AI technologies, smart services like chatbots can fulfill personal needs for Muslim travelers such as presenting a preferable TV show to watch or suggesting a Halal restaurant [16].

A vital component of the provision of goods and services is the digital platform. AI applications would provide Muslim travelers with a modern digital platform that is tailored for everything. New Halal ventures that do not take full advantage of social media's real-time engagement are losing innovative opportunities to reach the entire young Muslim community [60–63]. Thus, new AI applications must create tools to enhance interaction in order to meet these digital natives. Moreover, a digital platform is essential for Umrah and Haj. To fully offer outstanding service experience for customers, Umrah service providers need to step beyond offering basic education and logistics support. Data-driven services would be their medium of differentiating themselves. In Umrah travel, there are many service moments of truth for them to offer customized and pleasant interaction. Furthermore, virtual reality and augmented reality using AI applications could be used to increase the Muslim tourist experience during Haj and Umarah, particularly for those who experience first-time pilgrims.

VR may be a helpful tool for discovering Muslim heritage and visiting a number of Muslim tourist sites digitally. Neither tourism operators nor tourism service providers dealing with Muslim travelers are sufficiently covered these kinds of services. One of the promising areas that needs to receive more attention from Halal entrepreneurs using AI apps is virtual Halal tourism. Moreover, in travel experiences including interplanetary journeys, fantasy world tours, sports events, and large theme parks, VR plays a crucial role [64]. New technologies also improve accessibility for tourism and contribute to the preservation of heritage [65]. In addition, the status of visual reality in tourism has not been mapped yet [66].

3.3 Evaluation Phase

After the tourists finish their trip (post-trip stage), they usually evaluate the whole journey, leading to satisfaction and repeat visit intention. In this phase, valuable data can be gathered from chatbots, which could be analyzed by data scientists to

provide personalized services [67], make a prediction based on historical feedback data [2], help to share the whole experience, such as collecting reviews about the journey and planning for more future trips [12]. Additionally, tourists' satisfaction results from evaluating the services consumed in the trip, which is expressed in emotional response [68].

In this stage, traveling to and coming back from a destination, tourists face navigation and wayfinding difficulties, choosing a transport mode, and transit experience (e.g., in airports). This is an important stage to secure service providers and authorities and balance travel facilitation and identity verification as tourist's cross borders. Furthermore, AI can also be used in customer feedback, as both ML and NLP can be applied to make review analytics more effective [2].

In this stage, for Muslim travelers, digital travel companies and chatbots on smartphones play a critical role in helping with information for navigation and wayfinding. Therefore, it is required to encourage positive evaluation using AI applications, as people typically establish their first impressions when confronting others, and positive first impressions often lead to positive evaluation [69]. Supposing a tourist shows any negative impression on social media platforms. The AI tool analyses the tourist's intention and the context to automatically reach out with real-time interventions that are most likely to deliver a positive impact [2]. These interventions could range from providing additional information, and helping the tourist understand the situation to more options to meet the tourist's requirements [2].

4 Conclusion and Expected Implications

The theory of tourism consumption system (TCS) presented by Woodside and Dubelaar [70] stated that the consumption phase of tourism is an entire experience that involves gathering pre-trip, during the trip, and subsequent trip ideas, choices, and actions. The experiences that start with obtaining and collecting data and then create an impression of the destination that leads to expectations are called a pre-trip experience [71]. Woodside and Dubelaar [70] stated that as an accumulated phase, the entire tourist experience means that the experience that occurs at one point will impact the next level's experience. In the Halal tourism industry, the specific needs of this segment should be addressed by destination marketers who target Muslim tourists as an important segment of the global travel market. AI applications help in this matter to form Halal automation tourism experience. At the pre-trip point, intelligent automation can be used to provide travel inspiration to Muslim tourists and aid them in the processes of knowledge search, booking, and pre-arrival experiences. Moreover, AI will assist in understanding the needs of Muslim visitors and solve the problems facing Halal tourism organizations. Therefore, it is important to discover intelligent solutions to enhance Muslim tourists' experience at each point of their journey, from planning to sharing their travel feedback [72].

AI offers powerful solutions for Muslim travelers, such as social media analytics and tourism recommendation framework for information overload. Tourists can also obtain immediate data and assistants through AI-powered tools such as chatbots. Chatbots can help Muslim travel recognize several things relevant to Islamic travel planning, such as Halal airlines, Halal food availability, prayer times, and booking Halal entertainment. Furthermore, using AI in a hotel room can only suggest very few restaurant choices and interesting Islamic tourism sites on the basis of user profile data and past interests, such as places to visit and eat, things to buy, and people to meet. Moreover, AI applications might provide Muslim travelers with a modern digital platform that is tailored to his/her travel journey. It is recommended that social media's real-time engagement be utilized entirely to target the entire young Muslim community. Therefore, new AI applications must generate content that excites and ignites interaction in order to meet these digital natives. This may include immersive learning resources to teach Muslim pilgrims how to properly perform the Umrah and Haj rituals and to share their experiences with them.

AI could provide many socio-economic benefits in Halal tourism which could be sources of additional company turnover, jobs and government revenues. For example, it is essential to explore further innovations in the coming years and provide alternatives to financing and transacting for Muslim-friendly trips. Digital payments and Fintech firms are expected to continue contributing to global innovation in payment solutions [73]. Halal payments using AI, however, will play a role in taking advantage of this region as a creative opportunity. Moreover, 3D printing in Halal tourism field is not completely exploited as a creative opportunity. AI applications could be used to engage tourists to create their souvenirs in Halal space to increase the smart experience for Muslim travelers. In addition, new emerging technologies such as chatbots, virtual reality, language translators, etc., can be implemented effectively in Muslim-friendly hotels during the stay.

Moreover, supposing a Muslim tourist expresses a dissatisfaction on social media. In that case, the AI tool analyzes the tourist's intention and the context to reach out automatically with real-time interventions that are most likely to have a positive effect. Finally, this conceptual paper tried to fill the gap relevant to AI research in Halal tourism, which is very limited. For future research, developing a theoretical framework relevant to AI in Halal tourism industry is needed. Perhaps qualitative research accompanied by quantitative research is recommended to establish a new framework.

References

1. Addo E, Yagci K (2014) Automation, tourism. In: Encyclopedia of tourism. Springer, Cham
2. Gajdošík T, Marciš M (2019) Artificial intelligence tools for smart tourism development. In: Computer science on-line conference, Springer, pp 392–402
3. Battour M, Rahman MK, Rana MS (2019) The impact of PHTPS on trip quality, trip value, satisfaction and word of mouth. J Islamic Mark
4. Battour M (2019) Halal tourism: achieving Muslim tourists' satisfaction and loyalty

5. Battour M, Ismail MN, Battor M, Awais M (2017) Islamic tourism: an empirical examination of travel motivation and satisfaction in Malaysia. Curr Issue Tour 20(1):50–67. https://doi.org/10.1080/13683500.2014.965665
6. Tussyadiah I, Miller G (2019) Perceived impacts of artificial intelligence and responses to positive behaviour change intervention. In: Information and communication technologies in tourism 2019, pp 359–370. Springer
7. Ivanov S, Webster C (2019) Perceived appropriateness and intention to use service robots in tourism. In: Information and communication technologies in tourism 2019, pp 237–248. Springer
8. Murphy J, Gretzel U, Pesonen J (2019) Marketing robot services in hospitality and tourism: the role of anthropomorphism. J Travel Tour Mark 36(7):784–795
9. Murphy J, Hofacker C, Gretzel U (2017) Dawning of the age of robots in hospitality and tourism: challenges for teaching and research. Eur J Tourism Res 15(2017):104–111
10. Tung VWS, Law R (2017) The potential for tourism and hospitality experience research in human-robot interactions. Int J Contemp Hospitality Manage
11. Yeoman I, Mars M (2012) Robots, men and sex tourism. Futures 44(4):365–371
12. Tussyadiah I (2020) A review of research into automation in tourism: launching the annals of tourism research curated collection on artificial intelligence and robotics in tourism. Ann Tourism Res 81:102883. https://doi.org/10.1016/j.annals.2020.102883
13. Pinheiro AB, Pinto AS, Abreu A, Costa E, Borges I (2021) The impact of artificial intelligence on the tourism industry: a systematic review. In: de Carvalho JV, Rocha Á, Liberato P, Peña A (eds) Advances in tourism, technology and systems, pp 458–469. Springer, Singapore
14. Tussyadiah IP (2014) Toward a theoretical foundation for experience design in tourism. J Travel Res 53(5):543–564
15. Oecd K (2018) OECD science, technology and innovation Outlook 2018. OECD Publishing
16. Kabadayi S, Ali F, Choi H, Joosten H, Lu C (2019) Smart service experience in hospitality and tourism services. J Serv Manage
17. Samala N, Katkam Bharath S, Bellamkonda Raja S, Rodriguez Raul V (2020) Impact of AI and robotics in the tourism sector: a critical insight. J Tourism Futures. https://doi.org/10.1108/jtf-07-2019-0065
18. Ivanov SH, Webster C, Berezina K (2017) Adoption of robots and service automation by tourism and hospitality companies. Revista Turismo Desenvolvimento 27(28):1501–1517
19. Murphy J, Gretzel U, Hofacker C (2017) Service robots in hospitality and tourism: investigating anthropomorphism. In: 15th APacCHRIE conference, vol 31
20. Atrill P, McLaney EJ (2008) Accounting and finance for non-specialists. Pearson Education
21. Ivanov S, Webster C (2019) Economic fundamentals of the use of robots, artificial intelligence, and service automation in travel, Tourism, and Hospitality. In: Stanislav I, Craig W (eds) Robots, artificial intelligence, and service automation in travel, tourism and hospitality, pp 39–55. Emerald Publishing Limited
22. Kaartemo V, Helkkula A (2018) A systematic review of artificial intelligence and robots in value co-creation: current status and future research avenues. J Creating Value 4(2):211–228
23. Paschen U, Pitt C, Kietzmann J (2020) Artificial intelligence: building blocks and an innovation typology. Bus Horiz 63(2):147–155
24. Castillo-Montoya M (2016) Preparing for interview research: the interview protocol refinement framework. Qual Report 21(5):811–831
25. Samara D, Magnisalis I, Peristeras V (2020) Artificial intelligence and big data in tourism: a systematic literature review. J Hospitality Tourism Technol
26. Battour M, Ismail MN, Battor M (2010) Toward a Halal tourism market. Tour Anal 15 (4):461–470
27. Ryan C (2016) Halal tourism. Tourism Manage Perspect 19:121–123. https://doi.org/10.1016/j.tmp.2015.12.014
28. Henderson J (2009) Islamic tourism reviewed. Tourism Recreation Res 34(2):207–211. https://doi.org/10.1080/02508281.2009.11081594

29. Abror A, Patrisia D, Trinanda O, Omar MW, Wardi Y (2020) Antecedents of word of mouth in Muslim-friendly tourism marketing: the role of religiosity. J Islamic Mark
30. Yagmur Y, Ehtiyar R, Aksu A (2019) Evaluation of Halal tourism in terms of bibliometric characteristics. J Islamic Mark
31. Han H, Al-Ansi A, Olya HG, Kim W (2019) Exploring Halal-friendly destination attributes in South Korea: perceptions and behaviors of Muslim travelers toward a non-Muslim destination. Tour Manage 71:151–164
32. Battour M, Salaheldeen M, Mady K (2021) Halal tourism: exploring innovative marketing opportunities for entrepreneurs. J Islamic Mark
33. Battour MMK (2011) The impact of Islamic attributes of destination on tourists' motivation, satisfaction and destination loyalty. University of Malaya
34. Battour M (2017) Halal tourism and its impact on non-Muslim tourists' perception, trip quality and trip value. Int J Cult, Tourism Hospitality Res 2:1–24
35. Battour M, Ismail MN (2016) Halal tourism: concepts, practises, challenges and future. Tourism Manage Perspect 19:150–154
36. Al-Ansi A, Han H (2019) Role of Halal-friendly destination performances, value, satisfaction, and trust in generating destination image and loyalty. J Destin Mark Manag 13:51–60
37. Rashid NRNA, Wangbenmad C, Mansor KA (2020) Halal tourism: lessons for destination managers of non-Muslim majority countries. Int J Bus Soc 21(1):473–490
38. Battour M, Ismail MN, Battor M (2011) The impact of destination attributes on Muslim tourist's choice. Int J Tour Res 13(6):527–540
39. El-Gohary H (2016) Halal tourism, is it really Halal? Tourism Manage Perspect 19:124–130
40. Boğan E, Sarıışık M (2019) Halal tourism: conceptual and practical challenges. J Islamic Mark
41. Battour M (2018) Muslim travel behavior in halal tourism. In: Mobilities, tourism and travel behavior-contexts and boundaries: InTech
42. Battour MM, Battor MM, Ismail M (2012) The mediating role of tourist satisfaction: a study of Muslim tourists in Malaysia. J Travel Tour Mark 29(3):279–297
43. Battour M, Ismail MN (2014) The role of destination attributes in Islamic tourism. In: SHS web of conferences, vol 12, p 01077. EDP Sciences
44. Battour M, Battor M, Bhatti MA (2014) Islamic attributes of destination: construct development and measurement validation, and their impact on tourist satisfaction. Int J Tour Res 16(6):556–564
45. Vargas-Sánchez A, Moral-Moral M (2019) Halal tourism: state of the art. Tourism Rev
46. Ferreira DS, Ramos FRS, Teixeira E (2021) Mobile application for the educational praxis of nurses in the family health strategy: ideation and prototyping. Escola Anna Nery 25(1)
47. Kbaier MEBH, Masri H, Krichen S (2017) A personalized hybrid tourism recommender system. In: 2017 IEEE/ACS 14th international conference on computer systems and applications (AICCSA). IEEE, pp 244–250
48. Nilashi M, Ibrahim O, Ahmadi H, Shahmoradi L (2017) A knowledge-based system for breast cancer classification using fuzzy logic method. Telematics Inform 34(4):133–144
49. Ivanov SH (2019) Ultimate transformation: How will automation technologies disrupt the travel, tourism and hospitality industries? Zeitschrift für Tourismuswissenschaft 11(1)
50. Nunavath V, Goodwin M (2018) The role of artificial intelligence in social media big data analytics for disaster management-initial results of a systematic literature review. In: 2018 5th International conference on information and communication technologies for disaster management (ICT-DM). IEEE, pp 1–4
51. Sivarajah U, Irani Z, Gupta S, Mahroof K (2020) Role of big data and social media analytics for business to business sustainability: a participatory web context. Ind Mark Manage 86:163–179
52. Gavalas D, Konstantopoulos C, Mastakas K, Pantziou G (2014) Mobile recommender systems in tourism. J Netw Comput Appl 39:319–333
53. Pillai R, Sivathanu B (2020) Adoption of AI-based chatbots for hospitality and tourism. Int J Contemp Hospitality Manage

54. Melián-González S, Gutiérrez-Taño D, Bulchand-Gidumal J (2019) Predicting the intentions to use chatbots for travel and tourism. Curr Issues Tourism 1–19
55. Ukpabi D, Karjaluoto H, Olaleye S, Mogaji E (2019)Influence of offline activities and customer value creation on online travel community continuance usage intention. In: Information and communication technologies in tourism 2019, pp 450–460. Springer
56. Hsu P, Zhao J, Liao K, Liu T, Wang C (2017) AllergyBot: a Chatbot technology intervention for young adults with food allergies dining out. In: Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems, pp 74–79
57. Ukpabi DC, Aslam B, Karjaluoto H (2019) Chatbot adoption in tourism services: a conceptual exploration. In: Robots, artificial intelligence, and service automation in travel, tourism and hospitality, pp 105–121. Emerald Publishing Limited
58. Ukpabi D, Karjaluoto H, Olaleye SA, Mogaji E (2018) Dual perspectives on the role of artificially intelligent robotic virtual agents in the tourism, travel and hospitality industries. In: EuroMed academy of business conference book of proceedings. EuroMed Press
59. Neuhofer B, Buhalis D (2014) Issues, challenges and trends of technology enhanced tourism experience. The Routledge Handbook of Tourism Marketing
60. Situmorang SH, Rini ES, Muda I (2017) Customer experience, net emotional value and net promoter score on muslim middle class women in medan. Int J Econ Res 14(20):269–283
61. Sarwar A, Haque A, Yasmin F (2013) The usage of social network as a marketing tool: Malaysian Muslim consumers' perspective. Int J Acad Res Econ Manage Sci 2(1):93
62. Fadahunsi A, Kargwell S (2015) Social media, consumer behavior and marketing strategy: implications of "Halal" on Islamic marketing operations. J Small Bus Entrepreneurship Dev 3 (1):36–43
63. Idris J, Rahman SSA (2017) Aida model, social media marketing and decision to patronize of Muslim friendly hotel. J Manage Muamalah 7(2):3–15
64. Kim MJ, Lee C-K, Jung T (2020) Exploring consumer behavior in virtual reality tourism using an extended stimulus-organism-response model. J Travel Res 59(1):69–89
65. Bogicevic V, Seo S, Kandampully JA, Liu SQ, Rudd NA (2019) Virtual reality presence as a preamble of tourism experience: the role of mental imagery. Tour Manage 74:55–64
66. Yung R, Khoo-Lattimore C (2019) New realities: a systematic literature review on virtual reality and augmented reality in tourism research. Curr Issues Tourism 22(17):2056–2081. https://doi.org/10.1080/13683500.2017.1417359
67. Androutsopoulou A, Karacapilidis N, Loukis E, Charalabidis Y (2019) Transforming the communication between citizens and government through AI-guided chatbots. Government information quarterly 36(2):358–367. https://doi.org/10.1016/j.giq.2018.10.001
68. Nicolau JL, Más FJ (2008) Sequential choice behavior: going on vacation and type of destination. Tourism Manage 29(5):1023–1034.: https://doi.org/10.1016/j.tourman.2008.01.004
69. Tussyadiah IP, Park S (2018) When guests trust hosts for their words: host description and trust in sharing economy. Tour Manage 67:261–272
70. Woodside AG, Dubelaar C (2002) A general theory of tourism consumption systems: a conceptual framework and an empirical exploration. J Travel Res. 41(2):120–132. https://doi.org/10.1177/004728702237412
71. Wijaya S, King B, Morrison A, Nguyen T-H (2017) Destination encounters with local food: the experience of international visitors in Indonesia. Tourism Culture Commun 17(2):79–91. https://doi.org/10.3727/109830417X14966810027526
72. Battour M, Hakimian F, Ismail M, Boğan E (2018) The perception of non-Muslim tourists towards Halal tourism. J Islamic Mark
73. Ayedh A, Echchabi A, Battour M, Omar M (2020) Malaysian Muslim investors' behaviour towards the blockchain-based bitcoin cryptocurrency market. J Islamic Mark

Assistive Technology and Its Impact on Educational Achievement for Visually Impaired Students at SKPK Princess Elizabeth

Hussein Mohammed Esmail Abualrejal, Hassan Olayan Shtawi, Mohamad Ghozali Hassan, Amer Zaid Alqudah, and Amal Abdulwahab Alamrani

Abstract Assistive Technology is crucial for visually impaired students to gain access to information and educational learning. Students with vision problems face many difficulties in learning, and using assistive technology can help them meet their educational needs. This research aims to identify and describe the kinds of modern assistive technology accessible in the classroom, the modern technology available in school, the criteria used to determine Assistive Technology, and the challenges teachers and students face. An interview was used to collect data from five students and five teachers via a phone call interview at SKPK Princess Elizabeth. The study revealed that the utilization of current learning helps materials had a high commitment to understudies' relational abilities and schooling learning. In this manner, the learning helps materials emphatically sway on these students in their schooling regions and everyday life and accomplish adequate and well-rounded schooling quality for understudies with visual disability.

Keywords Assistive technology · Visual impairments · The achievement of education · SKPK Princess Elizabeth

H. M. E. Abualrejal · H. O. Shtawi (✉) · A. Z. Alqudah · A. A. Alamrani
Knowledge Science Research Lab, School of Technology Management and Logistics, College of Business, Universiti Utara Malaysia, Sintok, Kedah, Malaysia
e-mail: abualrejal@uum.edu.my

M. G. Hassan
Disaster Management Institute (DMI), School of Technology Management and Logistics, College of Business, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_69

1 Introduction

Assistive technology helps and inspires excellent blind learners to do their best work [1]. Those leading Assistant innovations are imaginative inventions that change learning for the sake of visual incapacity. Such development assists in identifying understudies who have a physical, substantial, or mental disability. This new creation provides teachers with creative resources to help understudies with special needs to solve their learning cycle problems. The goal of using assistive technology is to obtain as much guidance as required, regardless of location. Likewise, it can be said that modern tools encourage non-form beginners. In comparison, the learning environment for students with visual disabilities is minimal. Assistive technology has recently taken precedence; usability of innovation structures and publications is critical in this area. Assistive creativity provides learners with a variety of strategies, curiosity, newbie motivation, and cognitive load. Assistive Technology requires the use of some applications that support the learning cycle. Assistive Technologies-ology includes any item that enhances, retains, or revives devices [2].

To address their classroom limitation, students with a visual disability require assistive technology, such as reading apps, electronic Braille, and digital talking book devices [3]. It is difficult for them to receive proper educational knowledge, recreational activities, and social media experience if they cannot see well due to outdated assistive technology. Assistive Technology includes Braille typewriters and speech recognition systems. JAW has additional applications. They have access to cutting-edge information and communication technology, and assistive technology enables people with special needs to perform routine tasks [4]. However, it limits achievement for students with visual disabilities in SK Princess Elizabeth to recognize and describe modern assistive support practices that cater to students with particular needs. There is still a lack of access and usability to relevant assistive technologies [5], and there is a need for teachers and parents to be developed professionally. Therefore, this research aims to define and describe the various assistive technologies in the classroom, their availability in schools, and the problems teachers and students are likely to experience in Princess Elizabeth.

2 Literature Review

2.1 Assistive Technology

Assistive technology instruments include mechanical, electronic, and chip-based hardware, non-mechanical, non-electronic guidelines, and concentrated educational materials and processes for people with disabilities [6]. It also operates beyond the limits to encourage discovering and developing their learning from any perspective [7]. The other broad category of assistive technology equipment includes a wide range of devices from high to low tech that can be used for writing, reading, internet

access, communication, mobility, and leisure [8]. However, due to a lack of professional development or preparation for teachers, parents, and students, assistive technology is frequently misused or misapplied.

Technology is a backbone for organizational success in different industries like manufacturing [9–11], education organizations for teachers [12] and student [13], and Constructions industry [14]. In same time, the technology can cause information overload for students [15]. In case of students with visual disabilities must participate fully in guidance, which is frequently rich in visual content; therefore, assistant technology is one way to support them in this progression. These abilities enable understudies to have unrestricted access to a wide range of data [16].

2.2 Criteria for Selecting Assistive Technology

According to research, technological propels have broadened the correspondence between the externally weakened by touch, discourse, and remaining vision [17]. There is a variety of equipment and programs available for visually impaired and externally impaired students. They have been surveyed and used in their schooling in the nations of development. They introduce the use of computers adapted to speech using combined discourse [16]. Emphasized the importance of selecting the appropriate technology tool for students with disabilities. It is sometimes much more effective to search within the usual, current technologies for useful devices or software for students with disabilities rather than focusing solely on finding a specifically designed technology.

2.3 Teacher Challenges

The full potential of innovations cannot be fully realized due to a shortage of teachers with computer skills. To meet students' complex academic needs with physical disabilities in the classroom, special education teachers face various challenges. While some of these students were unable to read written texts, others had difficulty understanding spoken information. Some people find communication the most difficult challenge, while others find transitioning to classrooms to be the most difficult one. Teachers play a consistent and significant role in special education.

3 Research Methodology

A qualitative approach was used in this study for analysis. In this study, a blind school in Johor Bahru, at SKPK Princess Elizabeth, is chosen for a phone interview. The qualitative approach is the best method for this study because it is more

narrative about the school's actual situation that will be studied and interviewed, and because of the time constraints for this research. Qualitative data is evaluated based on what it reveals about the user's experiences, beliefs, and meanings and should not be used to fit pre-existing philosophies [18]. In this study, two types of data were used: primary and secondary. A telephone interview with five students and five teachers from SKPK Princess Elizabeth provided this study's primary data. This study's secondary data were gathered through a literature review, book references, and electronic media, such as blogs, posts, journal conference papers, and research-related studies.

3.1 Data Collection Method

Data for this study were gathered through telephone interviews with students at SKPK Princess Elizabeth, a special blind school in Johor Bahru. It was chosen for this research. The interview is a gateway to the qualitative data collection approach, which is commonly used in field studies [19]. Semi-structured interviews can provide accurate and comparable qualitative data [20]. Semi-led interviews were chosen to gather information in this investigation because the data provided by SKPK Princess Elizabeth in Johor Bahru, Malaysia on Assistant Technology and its impact on instructive accomplishment at SKPK Princess Elizabeth is valuable to educators and understudies in the business. The researcher had prepared a set of questions to serve as an interview guide to achieve the researcher's objectives. In this study, observation occurs when the researcher collects first-hand data or information from the interview process. The interview data were recorded in audio format.

3.2 Data Analysis Method

Inductive information understanding is a significant part of a subjective exploration. Gathering and assessing information is accomplished in persistent emphasis by conducting an advanced information assortment investigation [18]. The feelings, individual encounters, and biographies of the meeting should be reported in detail since they are significant information in the information investigation measure. The essential information is the appropriate responses given by an organization interviewee. Additionally, optional information was gathered from the writing audit of books and electronic media. The specialist would assess the information acquired and find the assistive innovation and its impact on instructive accomplishment.

4 Findings

4.1 Usage Assistive Technology in Teaching for Visual Impairments

Approaches to the use of assistive technology in education focus on using technology to practice or rehearsal and assist and promote learning. A text-to-speech screen reader may be used as a 'bridge' between the written text and the ability to understand the content aurally and cognitively, whereas a student with difficulties in sequencing text thoughts may use graphic design tools as a gateway to visual processing abilities. Thus, if assistive technology is successfully integrated into daily classrooms, students may have greater flexibility in performing activities that they have historically been unable to perform or have performed with significant difficulty; by necessary improvements or modified contact methods with others. To complete a task, technology is required.

The study has examined how much assistive technology was used in the classroom at SKPK Princess Elizabeth. All teachers (100%) confirmed using technology during the lesson. Moreover, a student who participated in this study agreed that a computer was available in every school lab. The use of assistive technology allows a visually disabled student to better adapt to regular academic learning (Fig. 1).

4.2 Assistive Technology Accessible in Classroom

According to the study's findings, each subject employs a variety of devices and materials during teaching and instruction. It is easy to understand that computer and information technology play an important role in business, industry, and education. In addition, Networking and information technology have become critical tools for academic success and participation in activities. This technology must be accessible

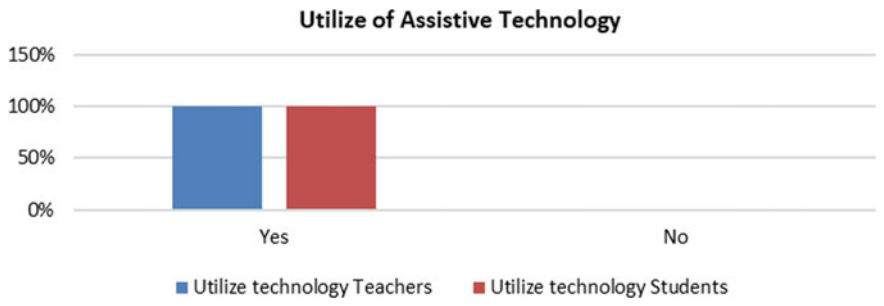


Fig. 1 Utilization of assistive technology by teachers and students

to all students, including those with disabilities. As teachers of vocational and technical education, we will use information technology to create and improve learning opportunities. The study sought to identify assistive technology in accessible classrooms for students with visual impairments.

The study's findings show that the use of Braille writer, large print textbook, and magnifying were the most instructional technologies used in SKPK Princess Elizabeth. According to the respondent, the devices mentioned above were used for specific subjects. They used tools, such as a talking calculator and a braille ruler for mathematics subjects. However, the students were more comfortable with those devices, especially with Braille.

Out of eighteen devices (80%), four were used during teaching and instructional lessons: Braille books, Abacus, maps and globes, and braille ruler. According to their subjects, some devices with 60% 3D materials: a talking calculator, a slate, and a stylus, were used. For visually impaired students, 3D materials were extremely useful. Students who were completely blind could touch and feel the 3D material to better understand certain topics. Furthermore, five out of the eighteen devices were rarely used: telescope and microscope, sound ball, tactual diagram, writing guide, and graphic kit. The optical device was the one that received the least attention. Only 10% of the equipment was used for sporting events (Fig. 2).

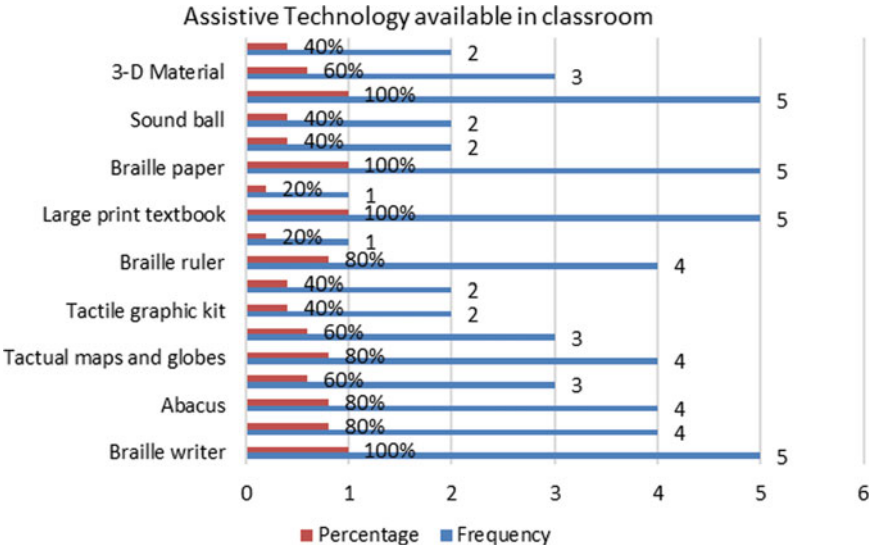


Fig. 2 Assistive technology accessible in a classroom

4.3 The Assistive Technology Used in SKPK Princess Elizabeth

The findings of the study revealed that some factors influenced educators, for example, the use of preparing advances, aptitude and abilities on use, time available for preparing, dependable force supply, acquisition, and organization consideration, given that the school provides the training innovation, the accessibility of rooms, and the adaptability of the innovation and its difficulty (Fig. 3).

The finding of this research shows that The Jaws program, Talking books, Braille, and Braille translation software were the most popular assistive technology accessible in SKPK Princess Elizabeth. Using Braille as the primary assistive technology in their studies was confirmed by 100% of the teachers who participated in this research via interview. It was also discovered that some teaching equipment used by teachers and students, such as the Dolphin pen, were used in one out of every five cases (20%).

4.4 Criteria Accessible Determine Assistive Technology

This study investigates the requirements for building up the available best assistive innovation suitable for SKPK Princess Elizabeth's outwardly disabled understudy. The respondents replied to the educator' inquiries.

The study discovered that it evaluated the acceptable Assistive Technology that fit students with visual impairments with varying needs. The term Assistive Technology refers to the use of technology to preserve, improve, and enhance visually impaired students' abilities.

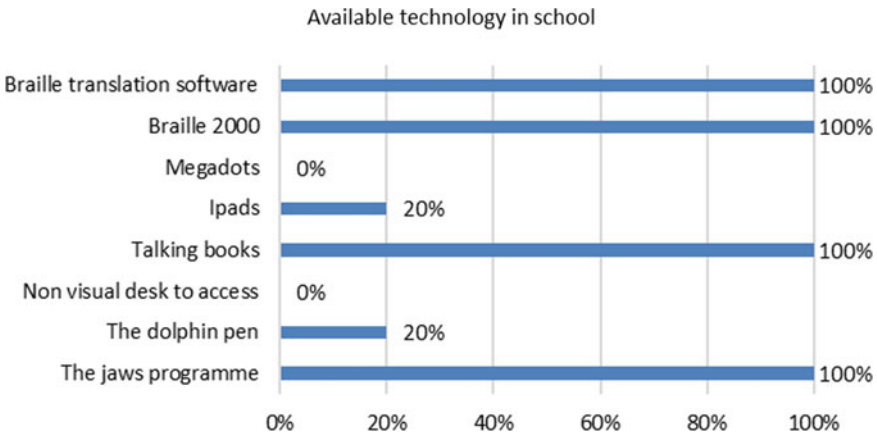


Fig. 3 Available technology in school

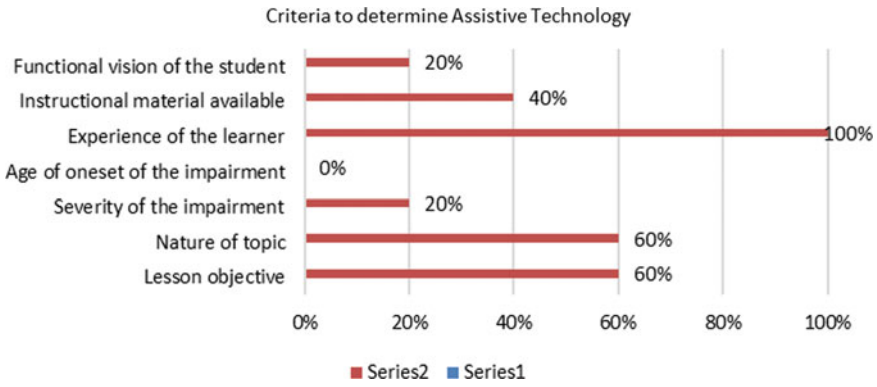


Fig. 4 Criteria to determine assistive technology

Figure 4 Five out of Five (100%) of the learning experiences represent an important factor that enables students with visual impairments to gain experience by observing various Assistant Technology programs offered by teachers. Furthermore, three out of five (60%) respondents stated that the subject matter's teaching goals and essence were very important when deciding on Assistant Technology (Fig. 4).

Furthermore, according to 40% of teachers surveyed, the availability of instructional resources was an important factor in selecting Assistant Technology. In addition, 20% of the severity of the disability and the student's work vision had to be considered when determining the required Assistant Technology for student educational needs.

4.5 The Challenges Faced by Teachers and Students with Visual Impairments Utilized Assistive Technology

Challenges faced by teachers. According to the report's findings, students' understandable levels differ; thus, teachers face challenges meeting their individual needs. Teachers who took part in the study claimed that learning aid resources, such as textbooks and modern assistive technology, were not as easily accessible as they should be and that the learning process was not as effective as it should be. Furthermore, the teachers who took part in the study discovered that students could not concentrate on the lesson, and their focus on the study was lost. Thus, the teachers were challenged to encourage these students and inspire them. Teachers face significant challenges in acquiring and using computers. The majority of responding teachers claimed that some students struggled with the English language, which hampered their efficient use of computers. Students who were completely blind lacked appropriate skills and required assistance from teachers in using

them on a regular basis in their instruction. Because a lack of vocabulary causes such difficulties among students, it may have a negative impact on their language competence. As a result, vocabulary is the most important aspect that learners face and struggle in learning.

Challenges faced by students. With 80% of respondents strongly disagreeing that computer lessons were not fun, and 20% of others denying computer lessons, it was decided that computer lessons were fun. Furthermore, 40% of students agreed that a lack of sufficient computers hampered computers' complete adoption. Access to students must be matched with the available equipment; a lack of equipment poses significant challenges for teachers; thus, students should have one-to-one computer access. In addition, some aspects of computer-attached devices, such as a mouse and keyboard, may not function properly when used by students. 40% agreed that students denied failure with the speech synthesizer. Professionals involved must have the knowledge and skills required to meet the needs of users. Competence refers to the availability of knowledge, skills, and experience needed to serve a client [21]. It necessitates additional advanced testing and preparations to ensure that equipment functions properly.

5 Recommendation

The school for students with visual impairments should be outfitted with modern devices that are less bulky, faster, easier to use and encourage independent learning. As a result, it is critical that assistive technology training be integrated into children's education with visual impairment for better adaptability and independence in society.

6 Conclusion

Education is a powerful tool for achieving economic growth in a country. Education is important for everyone, disabled or not. Assistive materials are necessary for the educational learning process because they make learning easier to understand and be more efficient for both students and teachers. These learning aid materials are especially important for students with disabilities. Learning aids materials, including assistive technology, are critical for these special students to improve their life. Visual impairments benefit the most from the use of assistive technology [22]. As a result of providing and protecting disabled students, their education and living standards can improve, and these students can also become skilled labour forces for the country. Many students cannot fully benefit from a traditional educational program because of a disability that limits their ability to participate in a typical classroom setting. For disabled students, it is critical to have someone who

understands and encourages them. The aptitudes and coordinate of educators are exceptionally critical for inabilities understudy in their school climate. The best guide content can permit understudies with a specific debilitation to approach the overall educational program. Individuals with disabilities use assistive technology to complete tasks that may be difficult or unusual, thereby increasing personal satisfaction. To expand society's adaptability and freedom, it is critical that training in assistive technology be combined with children's education with visual impairments.

References

1. Sarfaraz S, Muneer I, Liu H (2020) Combining fragment docking with graph theory to improve ligand docking for homology model structures. J Comput Aided Mol Des 34(12): 1237–1259
2. Silman F, Yaratan H, Karanfiller T (2017) Use of assistive technology for teaching-learning and administrative processes for the visually impaired people. Eurasia J Math, Sci Technol Educ 13(8):4805–4813
3. ERIC. https://eric.ed.gov/?q=technology+AND+Blue&ff1=locNorth+Carolina&id=ED584153. Last Accessed on 21 Jan 2021
4. Erdem R (2017) Students with special educational needs and assistive technologies: a literature review. Turkish Online J Educ Technol-TOJET 16(1):128–146
5. Inico JJ, Prabakaran TE (2018) An empirical study on the experiences of the visually-impaired students of loyola college with ICT support on using www.loyolacollege.edu. Int J Adv Res Comput Sci 9(1)
6. Blackhurst AE, Edyburn DL (2000) A brief history of special education technology. Special Educ Technol Pract 2(1):21–36
7. Desideri L, Di Santantonio A, Varrucciu N et al (2020) Assistive technology for cognition to support executive functions in autism: a scoping review. Adv Neurodev Disord 4:330–343
8. Wong ME, Cohen L (2011) School, family and other influences on assistive technology use: access and challenges for students with visual impairment in Singapore. Br J Vis Impair 29(2):130–144
9. Abu Al-Rejal HM, Udin ZM, Yusoff RZB (2018) The role of supply chain management toward performance of manufacturing companies in Yemen. Int J Supply Chain Manage 7(6): 92–99
10. Abu Al-Rejal HME, Udin Z, Hassan M, Sharif K, Al-Rahmi W, Al-kumaim N (2020) Green information technology adoption antecedence: a conceptual framework. In: Saeed F, Mohammed F, Gazem N (eds) Emerging trends in intelligent computing and informatics. IRICT 2019. Advances in intelligent systems and computing, vol 1073. Springer, Cham. https://doi.org/10.1007/978-3-030-33582-3_103
11. Makhloufi L, Al-Rejal HMEA, Mohtar S (2018) An investigation of the moderating effect of IT personnel capability on the relationship between intangible IT resources and IT infrastructure flexibility on the sustainable competitive advantages. J Asian Sci Res 8(9):277
12. Al-Rahmi AM, Ramin AK, Alamri MM, Al-Rahmi WM, Yahaya N, Abualrejal H, Al-Maatouk Q (2019) Evaluating the intended use of decision support system (DSS) via academic staff: an applying technology acceptance model (TAM). Int J Eng Adv Technol 8(6s3):565–571. https://doi.org/10.35940/ijeat.F1099.0986S319
13. Alamri MM, Al-Rahmi WM, Yahaya N, Al-Rahmi AM, Abualrejal H, Zeki AM, Al-Maatouk Q (2019) Towards adaptive e-learning among university students: by applying technology acceptance model (TAM). Int J Eng Adv Technol 8(6 Special Issue 3):270–276

14. Abualrejal HM, Udin ZM (2017) Green building toward construction sustainability: energy efficiency with material and design aspects. J Technol Oper Manage 100–109
15. Al-Kumaim NH, Hassan SH, Shabbir MS, Almazroi AA, Abu Al-Rejal HM (2021) Exploring the inescapable suffering among postgraduate researchers: information overload perceptions and implications for future research. Int J Inf Commun Technol Educ (IJICTE) 17(1):19–41
16. Okonji PE, Ogwezzy DC (2019) Awareness and barriers to adoption of assistive technologies among visually impaired people in Nigeria. Assist Technol 31(4):209–219
17. Alves CCDF, Monteiro GBM, Rabello S, Gasparetto MERF, Carvalho KMD (2009) Assistive technology applied to education of students with visual impairment. Rev Panam Salud Publica 26:148–152
18. Koch LC, Niesz T, McCarthy H (2014) Understanding and reporting qualitative research: an analytical review and recommendations for submitting authors. Rehabil Couns Bull 57(3): 131–143
19. Qu SQ, Dumay J (2011) The qualitative research interview. Qual Res Account Manag 8(3): 238–264
20. Newton N (2010) The use of semi-structured interviews in qualitative research: strengths and weaknesses. Exploring Qual Methods 1(1):1–11
21. Witte L, Steel E, Gupta S, Ramos VD, Roentgen U (2018) Assistive technology provision: towards an international framework for assuring availability and accessibility of affordable high-quality assistive technology. Disabil Rehabil Assist Technol 13(5):467–472
22. Ahmad FK (2015) Use of assistive technology in inclusive education: making room for diverse learning needs. Transcience 6(2):62–77

University Parcel Centre Services Quality and Users' Satisfaction in Higher Education Institutions: A Case of Universiti Utara Malaysia

Hussein Mohammed Abualrejal, Amer Zaid Alqudah, Ahmed Ali Atieh Ali, Oussama Saoula, and Tareq Khaled AlOrmuza

Abstract This study investigates the influence of the quality of service provided by the University Parcel Centre (UPC) on students' satisfaction. The UPC plays a significant function in offering key types of assistance to characterize, organize, and store the parcel, which shows up from the other couriers. It is also perceived as a one-stop parcel center point place where the users of Universiti Utara Malaysia can gather their parcels at UPC. This study is adopted from the SERVQUAL model. A quantitative method is applied, and a set questionnaire is developed using Google Forms and distributed to UUM students for data collection. A total of 152 respondents who are studying at UUM have been chosen for this study. Cronbach Alpha value was more than 0.7 value. The findings show that correlation from the tangible is ($r = 0.739$, $p < 0.00$), assurance ($r = 0.365$, $p < 0.00$) reliability ($r = 0.201$, $p < 0.01$), and empathy ($r = 0.228$, $p < 0.01$) which indicates influence on user's satisfaction while the responsiveness ($r = 0.114$, $p > 0.16$) did not have influence towards user' satisfaction. This research recommends extending the research scope by including more universities and factors to investigate the influences on users' satisfaction.

Keywords User satisfaction · University parcel centre · Service quality · Higher institutions · UUM

H. M. Abualrejal (✉) · A. Z. Alqudah · A. A. A. Ali · T. K. AlOrmuza
Knowledge Science Research Lab, School of Technology Management and Logistics, University Utara Malaysia, Sintok, Kedah, Malaysia
e-mail: abualrejal@uum.edu.my

O. Saoula
School of Business Management, Institute of Excellence for Islamicjerusalem Studies, Universiti Utara Malaysia, Sintok, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_70

1 Introduction

These days, quality has become a significant factor for thriving companies. Companies have a tough time satisfying and fulfilling the requirements and desires of their customers. User's satisfaction plays an important role based on service and product usage. In all industries, increasing level of service and satisfactions is crucial, organizations use related technologies like telehealth in healthcare [1] e-learning in education [2, 3], green technologies in manufacturing [4, 5]. E-commerce is a vital component of the impact on courier services that many users use the online system to order their items wherever they have a connection to the web [6]. Postal and postal businesses are the main operation for distributing orders, shipments, papers, letters, and written materials [7]. Therefore, courier services are essential operations for the exchanging of goods and information between users. The parcel collection systems in many universities have a considerable potential to satisfy students' demands and wishes to deliver in a timely and reliable manner in terms of their order package. Therefore, UniStorage is one of the organizations that oversees and controls parcel services in different universities. The organization plays a role in delivering mail and courier services by receiving, processing, storing, and distributing parcels to students and campus peoples [8]. Besides, the services provided by UniStorage on user demand can, as a result, understand the user's needs and wants. In relation, the UniStorage small division with the University Parcel Centre (UPC) name has recently been developed at UUM to provide students with access to their courier or parcel collection. UPC provides various services, such as mail and courier activities, payment counters, bus tickets, photocopies, and scanning [8]. Currently, UPC focuses mainly on the basic elements of self-collection and temporary housing of student parcels.

UPC charges lead to frustration for the customer because, in some third-party couriers and self-collection, the user has to pay for the delivery again. More importantly, the delivery price increases the parcel contentment period [9]. The package's incorrect layout makes it difficult for staff to locate the actual parcel, leading to longer delivery times. Lacking consistent awareness in the selection process, customers often face difficulties obtaining their parcels, and staff face difficulties in resolving consumer issues due to lack of information and experience in UPC services [10]. In UPC, users also disagree about parcel centres' role to meet and pick up parcels, particularly in higher quantities and bulkier parcels at UPC [11]. Based on the explanation above, the research aims at explaining and testing the relationship between the quality of service offered by UUM-UPC and student satisfaction level.

2 Literature Review

2.1 User Satisfaction

Satisfaction is described as the users' general approach reacting to the service provided [12]. The degree of user loyalty is calculated based on their prior

experience in the benefit obtained. User satisfaction will improve if the services they get satisfy their desires and standards [13]. Several indicators suggested, most of which are used by the researcher and generally for overall service quality, are service expectations and user satisfaction. Therefore, if the service quality is of a high standard, it will lead to continuous profitability [14].

2.2 Service Quality Dimensions

Service quality dimensions tangibility, reliability, responsiveness, assurance, and empathy play a significant role in customer satisfaction [15].

Tangibility. Tangibility refers to the appearance of hardware components, facilities, technology, and communication materials. Services cape is one of the researchers' words to demonstrate the spatial elements of the condition services. A variety of measurable items also cover, for example, circumstances, architectural architecture, and other facilities that can affect the degree of customer satisfaction. Encompassing disorder exists, such as the environment's scent, the vibration, and even the temperature. Even information overload could cause in communication [16], the service of tangible products can be defined as the external surroundings' practice that refers to information resources, the appearance of physical facilities, the private and any form of physical characteristics that the services offer.
H1: The tangibility and user satisfaction of using University Parcel Centre (UPC) services among UUM students is significantly related.

Assurance. According to [15], the guarantee element shows the capacity of intelligence and mutual understanding to gain user interest. The assurance shall provide a range of personnel qualities when communicating with the user with social principles of politeness and strong listening skills. The employees' best attitude and actions towards the user affect their satisfaction [17]. The staff has a full understanding and politeness to explain the concerns of the industry-based user. The staff providing the information called by users before improving their faith and belief in the service sector would continue to calculate the level of service in the courier industry [6]. Therefore, the parcel industry should provide extra resources and instruction for staff; they will improve their understanding of the operation. The reputation of the workforce reflects the belief that the employees communicate with the user competently. The parcel industry compares the personnel's position to the user's willingness to communicate efficiently with the parcel carrier and provide users with good knowledge of the facilities [18]. Be-sides, if any situations arise in the parcel industry, such as courier misplacement or loss of compassionate staff, the issues will be dealt with patiently.

H2: The assurance of University Parcel Centre (UPC) services and user satisfaction among UUM students is significantly related.

Reliability. Reliability is the willingness to deliver on the services as paid in terms of quality, regularity, and value as per the user's order. An indication of durability is the consumer's confidence in the service delivered, the reliable record, and the data's utility. Thus, it demonstrates that the dimension of reliability is essential [19]. Reliability also applies to staff's willingness to offer their services, which would impact the user trustfully and consistently [20].

H3: The reliability and user satisfaction of using University Parcel Centre (UPC) services among UUM students is significantly related.

Responsiveness. Responsiveness is the ability to assist the user by providing instant services without problems [15]. Besides, waiting time and the long line without a possible reason will create a meaningless negative impression of service quality. Punctual or prompt orders play a crucial role in the parcel arrival service to satisfy the user. Highly prospective parcel service element that instant ability to respond to the user will ascertain the level of performance of parcel service [21]. The parcel service industry has always paid massive attention to user satisfaction by implementing marketing-mixing strategies such as location, product, promotion, and price. The parcel shall always be punctual in its delivery to the user without error. Delaying the delivery and retrieval of parcels would have a direct negative effect on user loyalty.

H4: The responsiveness and user satisfaction of using University Parcel Centre (UPC) services among UUM students is significantly related.

Empathy. The empathy element aims to serve added concern and attention by offering services because of different users and different aspects [21]. Empathy underlines the need to fulfill personalized service via a range of times to help coordinate and understand the user's suggestion to ensure user loyalty while providing altered administrations to enhance user satisfaction [22]. Empathy has interconnections between the staff to acknowledge the staff's standpoint during interaction with the user professionally yields better outcome [23].

H5: The empathy and user satisfaction of using University Parcel Centre (UPC) services among UUM students is significantly related.

3 Research Methodology

This study depends on a quantitative approach to descriptive and hypothesis testing to evaluate the impact of service quality dimensions on satisfaction among UUM students on University Parcel Service (UPC). Data are gathered by a questionnaire adopted from previous research [24]. Meantime, descriptive design coordinates to determine the user satisfaction factor in any prospect to meet the UUM student

requirements. The inductive rationale for completing the study is based on empirical evidence [25]. This study used the deductive approach to analyze user satisfaction for UPC service among UUM students based on a formulated hypothesis. The questionnaire was generated via Google Forms and distributed to the respondents via Whatsapp for collecting data. A group sample of 152 students from University Utara Malaysia was randomly selected to answer this questionnaire. Then, the data were analyzed using the statistical tools of the Statistical Package for Social Sciences.

The data interpretation is a significant consideration for the researcher to check the precision, the descriptive statistics, the reliability test, and the correlation test. Descriptive statistics derive as a result of the overall set of questionnaires. Mean, median and mode are the three aspects to evaluate central trends. The mean arithmetic is the average and is used to measure dispersion from the mean. Reliability tests the response of the respondent to the overall item to measure the questionnaire. Based on the value of the items, independent measures of the same criteria will be correlated.

Cronbach's Alpha is used to calculate reliability problems and calculate the degree of user satisfaction by using UPC among UUM students. The higher the value, the better the instruments of measurement. Pearson Correlation Analysis is one of the statistical measurement methods used in this research to maintain the relationship between independent and dependent variables. If there is a correlation between two variables, both variables will be determined within the appropriate time frame. The numerical value of the correlation may be either positive or negative. The two-variable transform with the same order will be a positive correlation, while the variable transform to the opposite order will determine the correlation as negative.

4 Findings

The data were analyzed using the SPSS. Descriptive statistics, construct validity, reliability, and correlation analysis for data analysis were carried out in this study. Furthermore, demographic profiles include gender, age, race, semester, college, and route.

4.1 Descriptive Statistics

Figure 1 shows the sampling data of this research which is 152 respondents, who are the UPC service user that comes from various gender, age, race, semester, college, and route. According to the results, most of the sample who answered the questionnaire were females, making up 67.8% compared to the male representing 32.2%. The respondents' ages, ranging from 21 to 23 years old, make up 59.9%

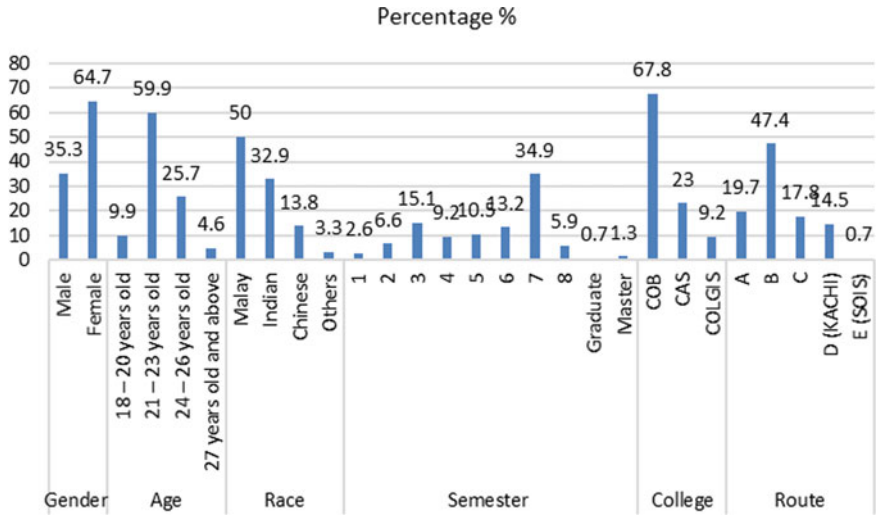


Fig. 1 Descriptive statistics. *Source* processed data

compared to the rest. In terms of race, Malay has the highest percentage of respondents, which conquered 76 percentage. As for the semesters, the highest percentage is from semester 7, making up 34.9%. The college of business (COB) holds the most elevated rate; it shows 67.8% compared to the rest. Lastly, most respondents from route (B) represent 47.4 percentage.

4.2 Validity and Reliability

Construct Validity. Factor analysis is employed to search patterns among the variables to determine if there is the underlying combination of the original factor that can summarize the original set [26].

Table 1 shows Kaiser-Meyer Olkin (KMO), which is the measurement of sampling adequacy full within the acceptable range. The value is significant when $p < 0.05$ for factor analysis. The Cronbach's Alpha is used to measure the consistency of responses across a set of questions (scale items) designed together to measure a particular concept [25]. Based on the results, Cronbach Alpha is more than 0.70. when it is ≥ 0.70 (sufficient/adequate); ≥ 0.80 (preferable), and ≥ 0.90 (important decisions on individual's fate) [27].

4.3 Means and Standard Deviations

Table 2 shows the descriptive statistic for the final list of variables. To make sure that easy interpretation is attainable. The result showed the mean value of

Table 1 Construct validity using factor analysis and Cronbach alpha

Variables	Items	Factor loading	Alpha	KMO	Barlett's test p-value	Eigen-value	% of variance explained
Tangible (TAN)	5	0.842–0.607	0.824	0.831	0.000	58.989	58.989
Assurance (ASU)	5	0.863–0.456	0.778	0.732	0.000	55.655	55.655
Reliability (REL)	5	0.908–0.845	0.923	0.864	0.000	76.629	76.629
Responsiveness (RES)	5	0.875–0.817	0.896	0.882	0.000	70.791	70.791
Empathy (EMP)	5	0.864–0.488	0.837	0.842	0.000	63.120	63.120
User satisfaction (US)	5	0.895–0.871	0.929	0.884	0.000	78.077	78.077

Table 2 Descriptive statistic of user satisfaction and service quality dimensions

Variables	Mean	Std. deviation
Tangibility (TAN)	3.4987	0.75888
Assurance (ASU)	4.0803	0.68755
Reliability (REL)	4.2487	0.82059
Responsiveness (RES)	4.2737	0.72849
Empathy (EMP)	4.1974	0.70004
User satisfaction (US)	3.4934	1.00724

Assurance, Reliability, Responsiveness, and Empathy with means of 4.08, 4.24, 4.27, and 4.19 respectively, which considered high means. Respondents agree that variables influence satisfaction. Furthermore, the respondents gave moderate answers to the remaining variables, for instance, tangibility and user satisfaction, with mean values that fall in the range of 3.50 and 3.49. The respondents have pointed to the medium level of satisfaction.

4.4 Pearson Correlation

Table 3 shows the correlation of both the independent variable (tangible, assurance, reliability, responsiveness, and empathy) and the dependent variable (user satisfaction). It shows a positive correlation in most variables except responsiveness. Tangibility has more influence on users satisfaction with the highest correlation ($r = 0.739$, $p < 0.01$) compared to another independent variable. However, responsiveness does not influence user satisfaction with a correlation ($r = 0.114$, $p > 0.05$).

Table 3 Correlation results of study variables

		Tangibility	Assurance	Reliability	Responsiveness	Empathy	User satisfaction
User satisfaction	Pearson correlation	0.739**	0.365**	0.201*	0.114	0.228**	1
	Sig (2-tailed)	0.000	0.000	0.013	0.163	0.005	
	N	152	152	152	152	152	152

*Correlation is significant at the 0.05 level (2-tailed). **Correlation is significant at the 0.01 level (2-tailed)

5 Recommendation for Future Research

Firstly, future research is required on the ongoing trends that evolve the technology. By this, the parcel service delivery can be handled smoothly. Building a parcel hub in each route will make it easier for the users. Parcel hub is a One-Stop Courier Service Centre specialized in packaging and courier services. On the contrary, the sample's scope is confined among UUM students only; future researchers have to widen their research scope to include more universities. Furthermore, future researchers adopt divergence underpinning theory, independent and dependent variables to carry out the same research project, but the data gathered will be based on a different perspective.

6 Discussion and Conclusion

This research intends to investigate the influence of the quality of service provided by the University Parcel Centre (UPC) on students satisfaction. According to mean values for all variables, respondents have agreed that all variables highly influence satisfaction except tangibility. The respondents have pointed to the medium level of satisfaction. Furthermore, the correlation result confirmed that tangible, assurance, reliability, and empathy significantly influence user satisfaction. In contrast, the relationship between responsiveness is not significant to user satisfaction. The result has accurately shown that four independent variables: tangible, assurance, reliability, and empathy have a p-value less than 0.05–0.01. Based on correlation results, this research support H1, H2, H3, H5.

However, the independent variable responsiveness has a p-value of more than 0.05, which does not support H4. In their diversion, the correlation analysis findings show that the independent variable, which is tangible, assurance, reliability, and empathy, is significant towards user satisfaction as a dependent variable, while for the aspect of responsiveness, it is not significant towards user satisfaction. To conclude, this research's findings are converse to the previous researchers, whereas the responsiveness implies relatively to user satisfaction. On the other hand, responsiveness should be emphasized more in future studies [21].

References

1. Al-Fadhli AA, Othman M, Al-Rejal HMA, Al-Jamrh BA (2018) Developing a telehealth readiness assessment framework and tool for least developed countries: a case of Yemen. Indian J Public Health Res Dev 9(11):1489–1489
2. Alamri MM, Al-Rahmi WM, Yahaya N, Al-Rahmi AM, Abualrejal H, Zeki AM, Al-Maatouk Q (2019) Towards adaptive e-learning among university students: by applying technology acceptance model (TAM). Int J Eng Adv Technol 8(6 Special Issue 3):270–276

3. Al-Rahmi AM, Ramin AK, Alamri MM, Al-Rahmi WM, Yahaya N, Abualrejal H, Al-Maatouk Q (2019) Evaluating the intended use of decision support system (DSS) via academic staff: an applying technology acceptance model (TAM). Int J Eng Adv Technol 8 (6s3):565–571. https://doi.org/10.35940/ijeat.F1099.0986S319
4. Makhloufi L, Al-Rejal HMEA, Mohtar S (2018) An investigation of the moderating effect of IT personnel capability on the relationship between intangible IT resources and IT infrastructure flexibility on the sustainable competitive advantages. J Asian Sci Res 8(9):277
5. Al-Rejal HMEA, Udin Z, Hassan M, Sharif K, Al-Rahmi W, Al-kumaim N (2020) Green information technology adoption antecedence: a conceptual framework. In: Saeed F, Mohammed F, Gazem N (eds) Emerging trends in intelligent computing and informatics. IRICT 2019. Advances in intelligent systems and computing, vol 1073. Springer, Cham. https://doi.org/10.1007/978-3-030-33582-3_103
6. Asogwa BE, Asadu BU, Ezema JU, Ugwu CI, Ugwuanyi FC (2014). Use of ServQUAL in the evaluation of service quality of academic libraries in developing countries. Libr Philos Pract 2014(1)
7. Noordin A, Hasnan N, Osman H (2012) Service innovation of postal and courier services in Malaysia: will it lead to customer responsiveness. Int Proc Econ Dev Res 42:205–209
8. UniStorageUUM (2020) https://uum.unistorage.com.my/UniStorage™, https://www.unistorage.com.my/index.php?route=common/home
9. Valaei N, Rezaei S, Shahijan MK (2016) CouQual: assessing overall service quality in courier service industry and the moderating impact of age, gender and ethnicity. Int J Manage Concepts Philos 9(2):144
10. Floridi L (2013) Information quality. Philos Technol 26(1):1–6
11. Kimelberg SM, Williams E (2013) Evaluating the importance of business location factors: the influence of facility type. Growth Chang 44(1):92–117
12. Mcdougall GHG, Levesque T (2000) Customer satisfaction with services: putting perceived value into the equation. J Serv Mark 14(5):392–410
13. George D, Mallery P (2003) SPSS for Windows step by step: a simple guide and reference. 11.0 update
14. Alqershi N, Ismail AI, Abualrejal H, Salahudin SN (2020) Competitive advantage achievement through customer relationship management dimensions. J Distrib Sci 18 (11):61–67(2020). https://doi.org/10.15722/jds.18.11.202011.61
15. Parasuraman A, Berry L, Zeithaml V (1991) A understanding customer expectations of service. MIT Sloan Manage Rev 32(3):39
16. Al-Kumaim NH, Hassan SH, Shabbir MS, Almazroi AA, Al-Rejal HMA (2021) Exploring the inescapable suffering among postgraduate researchers: information overload perceptions and implications for future research. Int J Inf Commun Technol Educ (IJICTE) 17(1):19–41
17. Neupane R, Devkota M (2017) Evaluation of the impacts of service quali-ty dimensions on patient/customer satisfaction: a study of private hospitals in Nepal. Int J Soc Sci Manage 4 (3):165–176
18. Prakoso AF, Nurul R, Wulandari A et al (2017) Reliabil-ity: responsiveness, assurance, empathy, and tangible: still can satisfy the customer ? Int J Bus Manage Invention 6(3):68–75
19. Karatepe OM, Yavas U, Babakus E (2005) Measuring service quality of banks: scale development and validation. J Retail Consum Serv 12(5):373–383
20. Saraei S, Amini AM (2012) A study of service quality in rural ICT renters of Iran by SERVQUAL. Telecommun Policy 36(7):571–578
21. Yee HL, Daud D (2011) Measuring customer satisfaction in the parcel service delivery: a pilot study in Malaysia. Bus Econ Res 1(1)
22. Meng Q, Zhou Q (2016) Assessment of express delivery service based on SERVQUAL model. In: Wuhan international conference on E-business. Association for Information Systems AIS Electronic Library (AISeL)
23. Bahadur W, Aziz S, Zulfiqar S (2018) Effect of employee empathy on customer satisfaction and loyalty during employee–customer interactions: the mediating role of customer affective commitment and perceived service quality. Cogent Bus Manage 5(1):1–21

24. Sesen H, Ekemen MA (2020) Data on personal and contextual factors of university students on their entrepreneurial intentions in some Turkish universities. Data Brief 28:105086
25. Saunders M, Lewis P, Thornhill A (2019) Chapter 4: understanding re-search philosophy and approaches to theory development. In: Research methods for business students (Issue March) (2019)
26. Cooper DR, Schindler PS, Sun J (2006) Business research methods, vol 9. Mcgraw-hill, New York, pp 1–744
27. DeVellis RF (2016) Scale development: theory and applications vol 26. Sage publications

Does the Capability of Information Management, Knowledge Management, Innovation and Technology Affect the Sustainable Competitive Advantage of SMEs in Malaysia?

Nurul Huda Md Yatim, Najatul Afiqah Mohd Affandi, and Hazila Timan

Abstract Over the past few years, the development of SME has grown tremendously. Taking into account the envisaged future study expansion of SMEs, this study aimed to assess the capabilities of information management, knowledge management, innovation and technology aspects. Capability is the ability that SMEs own in mobilizing their resources to achieve desired goals. However, is there a link how SMEs' capabilities affect the sustainable competitive advantage? More specifically, are SMEs more likely dependents on these capabilities to remain sustain? Therefore, this paper attempts to investigate the answers by compiling data from 210 owners of SMEs using purposive sampling. All relationships were found to be significant; only information management capability was found not to be significant.

Keywords Sustainable competitive advantage · Dynamic capability · RBV · SMEs · Malaysia

1 Introduction

In twenty-first century, and across many sectors, it is understood and should be noticeable that gaining competitive advantages are the most challenging issues faces by firms [1–4]. This issue become critical particularly for small medium enterprises. With highly competitive market environment, it is important for SMEs to remain sustainable in the market and lasted longer over period of time. It is important to know the percentage of firms that close their business in a given year

N. H. M. Yatim (✉)
Fakulti Pengurusan Maklumat, Universiti Teknologi MARA, Cawangan Johor, Kampus Segamat, Malaysia

N. A. M. Affandi · H. Timan
Fakulti Pengurusan Maklumat, Universiti Teknologi MARA, Puncak Perdana, Malaysia

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_71

and to compare it over time and across countries. As reported in SME Annual report 2015/2016 of Malaysia, within the first five years, more than half of SMEs fail to start-ups, while the rates of surviving firms varying from less than one in five firms in Lithuania to about two-thirds in Sweden. Meanwhile, the survival rates of start-ups are consistently higher in Austria, Belgium, Luxembourg, the Netherlands and Sweden compared in other countries [5]. It shows that SMEs need to have a strong shield since their capability to survive can avoid them fall into mortality conditions. Whatever happened in other countries should be learnt by Malaysian SMEs for betterment. In this context of study, SMEs of services sector become the main subject, as SMEs amongst sector that contributed toward the national GDP with 37.1% in 2019 as compared to 36.6% in previous year. Thus, affirming that SMEs have substantial roles toward Malaysia's economic growth.

Therefore, based on the existence of the value and abundance of capabilities created by the SMEs, this paper highlighted whether these capabilities are behind the factors that contributed to the sustainable competitive advantage (SCA) of SMEs. It provides a guideline for SMEs in providing a complete understanding on the importance to understand the capabilities of information management and knowledge management, especially for the stakeholders as a shred of admissible evidence as boosting the image of the organization due to rapidly evolving communication technologies thus can stay competitively and productively. In additions, SMEs that using web-based marketplace gain productivity up to 11%. This web-based entrepreneurship creates business players with creativity and innovations thus allowed them to gain competitive advantages. The idea of leveraging web-based technology to grow business had been widespread across global. The example of companies that leveraging the internet include Google, Facebook and Twitter [6].

Next, the use of technology has driven to the innovation development. The results of the study conducted by Ismail et al. [7] indicated that innovation is one of the essential factor for companies that desires to grow the business and become competitive in respect of their national and international competitors. This would provide opportunities for SMEs to sustaining their business and also gain market share in marketplace. Thus, their survival in Malaysian market will depend on their ability to innovate thus made them ready to face challenge from other SMEs. Therefore, next section onwards will discuss whether the capability of information management, knowledge management, innovation and technology affect the sustainable competitive advantage of SMEs in Malaysia or not.

2 Literature Review

2.1 Capabilities as Strategic Resources

Understanding the types of resources and capabilities possess are important for all SMEs which present them a competitive advantage and at the same time contribute to the SCA. Although in Malaysian SMEs, many resources have been produced to

achieve SCA, whether these resources have the capabilities to avoid becoming not valuable, rare, imperfectly imitable, and not substitutable is still unknown. SMEs must know what their resources are. Besides, they need to know how to use them and turn them into an advantage for their business. Information is one of the vital resources in a business organization. According to Barney [8] information as a strategic resource is one of the competitive advantages for an organization. One question that needs to be asked, however, is whether the information is really a strategic resource, especially in the IR4.0 era. Albeit many studies earlier focusing on the importance of information management, but people still not aware of the needs of it. As agreed by Rad et al. [9] better decision making, efficiency enhancement and allow organizations to gain competitive advantages [10] is part of the benefits gained from good and quality information. Rahimli [11] mentioned that business owners must develop their capabilities to adapt with the changes and learn quickly so that they may surpass their competitors with the ability to generate, achieve and distribute knowledge through the organization. The importance of distributing knowledge through the whole organization is that each part of the organization will understand the mission and vision to be achieved and can work together to generate the new idea for the organization improvement and makes the organization stand taller as compared to other competitors. Coplin [12], argues that business can be place in advantageous positions ahead of their competitors if they able to explore and utilizing their resources and skills for competitive advantages.

The KM capability meanwhile is dynamic in the sense that it permits firms to dynamically manage their knowledge bases by reconfiguring and realigning the processes of knowledge exploration, retention, and exploitation inside and outside the firm [13]. It shows that innovation and technology are important factors that may contribute to it. However, the implementation of those cannot be in place without information as a driver to support. The fourth industry is also known as Industry 4.0, represents a special challenge for businesses in general and SMEs specifically [14]. At present, the fourth industrial revolution (IR) becoming an important agenda to be discussed in all countries around the globe. Ready or not, people, especially in the industry including SMEs, must no longer walk, but run and drive into this tunnel to make sure their organization sustains and competitive enough to strive in the industry.

It is interesting to note that the ICT adoption among SMEs is no longer merely a choice, but it is a must in today's globalized world. It is showed that ICT plays a pivotal role in smoothening the business activities in SMEs, in order for SMEs to stay competitive and also changing consumer behaviour [15]. Nevertheless, the technology always acted as a key enabler in smoothening the organization. The increased use of information and communication technology (ICT) in businesses, including the internet can lead to higher productivity and efficiency in all aspects of the value chain.

Those issues discussed on SMEs remained obstacles and unsolved, thus providing this study with the impetus to further investigate and overcome the issues of sustainable competitive advantage (SCA) which must be achieved. Therefore, in order to overcome the issue, the capability of a company in producing VRIN

service is important for it to remain sustainable in the market. The aspects of the capabilities that will be discussed in this paper are information management, knowledge, innovation, and technology or "IKIT" towards SCA.

2.2 Sustainable Competitive Advantage

Sustainable competitive advantage (SCA) as asserted by Barney [8], is the long-term benefit of implementing some unique value, and create the strategy which competitors do not able to implement and duplicate it simultaneously, thus lead to inability to gain benefits from it. Kuncoro and Suriani [17] added that through SCA the company can provide products qualities that were improved according to the customer demands and complaints, customer needs, procurement of new markets and product innovation. This is where the value of SCA can be seen in that particular company. The same authors also mentioned other ways to see the achievement of SCA in a corporate is the products or services provided by the company cannot be duplicate or it will cost the competitors in a big amount. There are some indicators to measure SCA such as product uniqueness, product quality, and competitive price. As for this study, the researchers used definitions by Haseeb et al. [18], referring to the importance of strategic alignment in displaying the beneficial functions of social and technological factors on sustainable competitive advantage.

2.3 Theoretical Reviews

This study will use Resource-based view (RBV) theory and dynamic capability (DC) as underpinning theories from resources and capabilities. The theories were identified based on the literatures reviewed and framework proposed earlier by previous scholars as follow.

Resource Based View Theory. This paper mainly aims to propose the RBV of strategic management from a resource perspective. But the concept of valuable in terms of resources seem to be vague [19]. This is because, Barney did not clearly mention what kind resources more valuable compared to others as different organizations will have different value of asset. When talking about RBV, it is normally based on the value, rareness, imitability and non-substitutability [8]. However, according to Madhani [20], a resource is valuable when it is fully utilize and able to uplift the organization to improve its performance and stay competitive with their competitors. Otherwise, when the resource is exploited at worst, a competitive disadvantage may be gained because the resource is not valuable to the organization. In this context, the resources can only beneficial when the SMEs have certain capabilities. The important and contribution of information management undoubtfully important to ensure organizational success and effectiveness [9].

In order for the organization to sustain the knowledge, they need to manage the information systematically. Although knowledge management (KM) is significantly give impact to the performance of SMEs, but the needs of managing information is more crucial than managing knowledge since knowledge is acquired from experience in a process of learning, and experiences represent opportunities [21]. But opportunities only can be created with the availability of information that exploited the creation of new products and services [22]. However, little work in RBV has been made to grasp the role of information as the crucial source of competitive advantage, even has highlighted the importance of knowledge.

Dynamic Capability. In SMEs environments, firms possess, develops, creates routines and same products/services to the customers especially in services sector due to heterogeneity of business environments. Since it only focuses on the capabilities towards sustainable competitive advantage, besides of RBV, this study also used DC to support the variables. DC derived from RBV theory of the firm of Barney's that result of applying a company's capabilities (competencies) to effectively use and develop its heterogeneous and imperfectly immobile resources in order to achieve competitive advantage are the core competencies, while the RBV primarily addresses a firm's existing resources, the dynamic capabilities view emphasizes the reconfiguration of these resources [23, 24]. This also alligned with Wójcik [25], asserted that the capabilities perspective has evolved within the resource-based view (RBV) that focused on endogenous factors, for instance the interior of the company, its resources, competencies and capabilities.

Research Framework. In this context capabilities that have been studied has bridging the theory of RBV and DC into SCA. After considering all the capabilities as discussed by previous scholars, this paper integrates their works and develops a framework serves as a guide or direction to this study that contributes to the body of knowledge by taking RBV and DC as underpinning theories that mentioned information management, knowledge management, innovation and technology as part of capabilities towards sustaining the competitive advantage for SMEs in Malaysia (Fig. 1).

3 Methodology

This study was carried out in a quantitative manner and data was collected at one time using a set of questionnaires. The type of sampling used was purposive sampling. Owners of the SMEs from the service sector in Peninsular Malaysia were chosen as a population. A total of 1000 questionnaires were distributed and only 210 were usable for subsequent analysis with an effective response rate of 21%. It was concluded that the sample is representative of the population of interest. Statistical technique version 21 and SmartPLS version 3.2.8 were used in analyzing the findings for this research. Table 1 shows the response rate for this study.

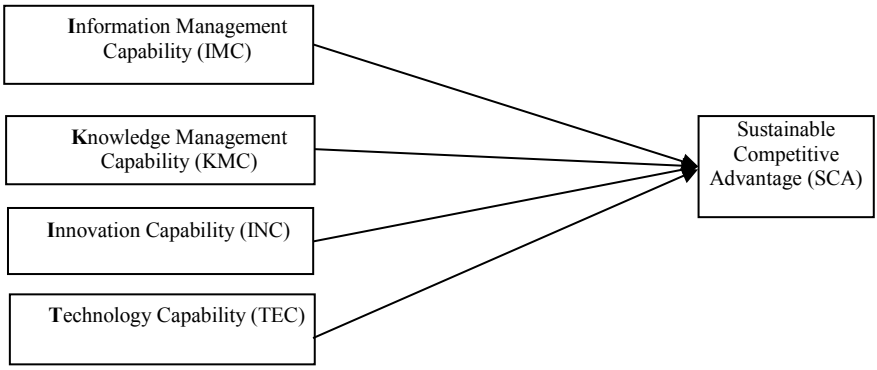


Fig. 1 A research framework

Table 1 Response rate

Response	Frequency/rate
Questionnaires distributed	1000
Questionnaires returned	252
Valid questionnaires	210
Response rate	25.2%
Response rate after data entry	21%

The skewness and kurtosis values for all variables in this study were also checked in order to verify that they did not veer from normal. It was found that the assumption of normality is achieved for all the variables with skewness ($\beta = 5.019$, $p < 1.570$) and Mardia's multivariate kurtosis ($\beta = 74.015$, $p < 0.01$), showed that the data collected was not multivariate normal. To remove doubts over the possibility of biasness in the cross-sectional survey, the common method variance using Harman's single factor test was performed. Factor one accounts for only 30% of the total variance. Thus, common method variance was not a serious threat in this study.

Subsequently, the measurement model was assessed for its reliability and validity prior to testing the proposed relationship between the variables. Reliability analysis was ascertained using composite reliability with loadings exceeding the cut-off value of 0.7 [26–28]. As for validity, content validity was performed during the pretesting of the questionnaire to measure the adequate coverage on the investigative questions guiding this study.

4 Findings and Analysis

The analysis for the structural model evaluation based on the effect size (f^2). Normally researchers evaluate the effect size using Cohen [29], who explained the relative impact of a predictor construct on an endogenous construct (dependent variable). Briefly, it assesses how exogenous construct (independent variable)

Table 2 Result of structural model

Hypothesis		Std beta	Std error	t-value	p-value	f^2	Results
H1	IMC → SCA	−0.046	0.052	0.868	0.193	0.004	Not sig
H2	KMC → SCA	0.082	0.075	1.092	0.138	0.007	Sig
H3	INC → SCA	0.467	0.077	6.072	0.000	0.208	Sig
H4	TEC → SCA	0.249	0.093	2.667	0.004	0.056	Sig

explains certain dependent variables in terms of R^2. According to Sullivan [30], effect size can refer to the raw difference between group means, or absolute effect size, as well as standardised measures of effect, which are calculated to transform the effect to an easily understood scale. The change in the R^2 values is calculated by estimating the PLS path model twice, with and without the inclusion of the exogenous latent variable. All relationships were found to be significant; only information management capability was found not to be significant.

Previous studies show a significant result on information management capability (IMC). However, the findings gathered for this research shows the contradict results which SCA was not positively influenced by IMC. In terms of Knowledge Management Capabilities, Innovation Capabilities and Technology Capabilities, the results show a positive relationship and direct effect towards sustainable competitive advantages. Table 2 displays the path coefficients, the *t*-values and their significance levels, *p*-values and effect size (f^2).

5 Discussion

5.1 Effect of Information Management Capabilities Towards Sustainable Competitive Advantages

The insignificant result of information management capabilities towards sustainable competitive advantages may be caused by most SMEs from service sector do not wanted to conceal their business information thus this has given a huge impact on the result. Since most of the service sector owner are the inheritors of their family business, all the information management capabilities are already made available for them to use. Therefore, they do not realize the true meaning and importance of information management capabilities towards their business growth. These inheritors also already get used to the current workflow and were used to follow the decision made by the dominators of the company, so they are just maintaining the business flow without any change for improvement.

The next reason for the insignificant result is that the complexity of information management capabilities itself. Migrating the way of information for a company is recorded from physical filing system to electronic filing system itself is not an easy and simple task. Added with the fact that old school business owner who are not an information technology (IT) savvy has contributed to bigger influences on this

issue. Most of the business owners have the same view which 'it is easier for the electronic records to be copied by others and most of electronic filing system are not that safe to use'. This issue has supported by Neirotti et al. [31] which mentioned that "*the limited effect of public funding measures can be in part explained by the lack of a deep understanding of how SMEs' needs and difficulties in using ICT change in response to their external environment*". By right, Tseng and Lee [32] suggested that they could have come with the alternative to provide incentive scheme in order to prevent the information from leaking to the competitors.

5.2 *Effect of Knowledge Management Capabilities Towards Sustainable Competitive Advantages*

Knowledge management capabilities shows a positive effect due to the ability of creating new knowledge and utilize the knowledge for the business' improvement even in a strong competitive environment. This statement was supported by Tseng and Lee [32] which mentioned that "*it is the role of the enterprise itself to effectively develop its knowledge management capability in order to provide and share intangible assets to win market competition*". Despite of many companies offer the same services; knowledge management capabilities has proved that how the owner of the knowledge utilize is what matters in keeping the company to stay in front compared to other competitors. Different experience contributed to the different knowledge utilization thus led to the different outcome for the company profit.

Chih et al. [33] pointed out that KM capability is an important factor that influences organizational performance. Since most of the SMEs were inherited from previous ancestors, most of the ideas and workflow has been decided and the current owner just have to follow the existing flow. However, most current owner should be given the opportunity to apply new ways of developing new products/services based on their own new knowledge. As a newbie, they might have new ideas and new problem-solving styles that can helps in increasing the company capital. The difference between tacit and explicit knowledge can be clearly seen at this position. Rahimli [11] mentioned that explicit knowledge is the knowledge that can be transferred through formal language while tacit knowledge is difficult to be transferred and can the implementation or utilization of tacit knowledge can only be done by the owner of the tacit knowledge themselves. This tacit knowledge are the ones that play important role towards the competitive intelligence in the company.

5.3 *Effect of Innovation Towards Sustainable Competitive Advantages*

Kuncoro and Suriani [17] refers to innovation as the ability to observe consumers in order to create satisfied customers by offering new products and the action were

followed by creating innovation as business strategies to overcome competitors and to meet the market demand. The result of this study also has shown a positive relationship of innovations towards the sustainable competitive advantage of SMEs in Malaysia. A lot of people have affected by the pandemic and lots of them has started to develop their small business. Thus, this has made the competition between each business owner becomes higher from day to day. As the results, the business owners must come out with the brand-new ideas and strategies; either implicit or explicit, to make their business at least keep floating rather than sinking.

There are a few clear examples that can be seen in how the SMEs push themselves to innovate their businesses such as the variation of flavour for the same food products, the use of digital marketing strategies and the use of raw materials from lower cost but able to provide the same product qualities. That were among many other examples that can be seen from all the SMEs struggles in order to keep their business alive. Quaye and Mensah [34] mentioned that there are a handful of innovation that can be considered by the business owners for instance marketing innovation, pricing innovation, promotion innovation and retail innovation. All they have to do is to keep on gathering knowledge so that they are ready to grab the opportunities to become the first inventors for their products and services.

5.4 Effect of Technology Towards Sustainable Competitive Advantages

Technology also has shown a positive influence towards sustainable competitive advantages of SMEs in Malaysia. As we can see, most people have started to go cashless while enjoying their shopping. Besides, some of the customers opt for online shopping especially during movement control order (MCO) due to pandemic Covid-19. This is where people started to familiarize themselves with Shoppe aps, Lazada aps, Grab food, Food Panda and other business application that available through the mobile phones. This has been proven that technology has become an upper hand for the business development.

According to Haseeb et al. [18] in order to be able to succeed in a highly competitive environment, SMEs must have the ability to adopt with these new used of technologies. The authors also indicated that SMEs must ensure better information technology (IT) skills and implementation as both of these key points can helps in increasing the sustainable competitive advantages. It is undeniable that most of the business owners has started grab the benefits from the technology advancement. This statement was supported by Quaye and Mensah [34] which mentioned that those business owners has started to develop websites and other social network sites such as Facebook, Instagram and Twitter so that they can provide service and product information to their customers. Besides they also use these sites to communicate and gather their customers' feedback for the products and services improvement.

6 Conclusion

When this study decided to measure the SCA in Malaysian SMEs, the relationship between the capabilities of information management, knowledge management, innovation and technology are crucial. RBV is a well-known theory among researchers, while DC is derived from RBV. Meanwhile, the propose of new framework of SCA in this paper is not the matter of adding a new concept, but it is more on creating the alertness to all SMEs to develop a good information management since the GDP of Malaysia is highly from SMEs. Although all the capabilities seem to be important among SMEs, but the existence and awareness among it seems vague, especially there is no study before that have been combining all four variables together.

Acknowledgements This work is a part of a project submitted to Universiti Teknologi MARA, Cawangan Johor, Kampus Segamat, Malaysia. The gratitude also goes to the anonymous researchers whom their articles have been cited in this paper.

References

1. Eniola AA, Ektebang H (2014) SME firms performance in Nigeria: competitive advantage and its impact. Int J Res Stud Manage 3(2):75–86
2. Munir A, Lim MK, Knight L (2011) Sustaining competitive advantage in SMEs. Procedia Soc Behav Sci 25:408–412
3. Supriyadi (2017) Analysis effect of resources and dynamic capabilities to sustainable competitive advantages and it's implications to the firms performance. Int Bus Res 10(10):209
4. Tasmin R, Takala J, Bakr AAA, Shylina D, Nizialek I, Rusuli MSC (2016) Sustainable competitive advantage (SCA) analysis of furniture manufacturers in malaysia: normalized scaled critical factor index (NSCFI) approach. Manage Prod Eng Rev 7(2):73–85
5. SME Corp (2018) Malaysia: SME annual report 2017/2018, Kuala Lumpur
6. Malaysian Ministry of Finance (2014) Part 3: economic performance and prospects. Economic report 2013/2014
7. Ismail K, Zaidi W, Omar W, Soehod K, Senin AA, Akhtar CS (2010) Role of innovation in SMEs performance: a case of Malaysian SMEs. Math Methods Eng Econ Role 145–149
8. Barney J (1991) Firm resources and sustained competitive advantage. J Manage 17(1):99–120
9. Rad AM, Shams G, Naderi B (2009) Information management and the role of information and knowledge managers: managers' perception. In: Proceedings of EDULEARN09 conference, Barcelona, Spain, pp 2504–2516
10. Porter ME, Millar VE (1985) How information gives you competitive advantage. Harvard Buiness Rev 63(4):149
11. Rahimli A (2012) Knowledge management and competitive advantage. Inf Knowl Manage 2 (7):37–43
12. Coplin LCH (2002) Competitive advantage and the SMEs: the role of distinctive competences as determinants of success, are there differences across gender, sector and size? Doctoral dissertation, Universitat Autonoma De Barcelona
13. Lichtenthaler U, Lichtenthaler E (2009) A capability-based framework for open innovation: complementing absorptive capacity. J Manage Stud 46(8):1315–1338
14. Sommer L (2015) Industrial revolution—industry 4.0: are german manufacturing SMEs the first victims of this revolution? J Ind Eng Manage 8(5):1512–1532

15. Economic Planning Unit (2015) Eleventh Malaysia plan strategy paper 18: transforming services sector, Putrajaya
16. Zeng F, Lee SHN, Lo CKY (2020) The role of information systems in the sustainable development of enterprises: a systematic literature network analysis. Sustainability 12(8):3337
17. Kuncoro W, Suriani WO (2018) Achieving sustainable competitive advantage through product innovation and market driving. Asia Pac Manage Rev 23(3):186–192
18. Haseeb M, Hussain HI, Kot S, Androniceanu A, Jermsittiparsert K (2019) Role of social and technological challenges in achieving a sustainable competitive advantage and sustainable business performance. Sustainability 11(14):3811
19. Wade M, Hulland J (2004) Review: the resource-based view and information systems research: review, extension, and suggestions for future reserach. MIS Q 28(1):107–142
20. Madhani PM (2010) The resource-based view (RBV): issues and perspectives. J Res Prestige Inst Manage 1(1):43–55
21. Unger JM, Homburg B (2006) Entrepreneurial success: the role of human capital and learning. Unpublished doctoral dissertation, Justus-Liebig Universität, Gießen, Germany
22. Alicia E, Blaga L (2015) Entrepreneurship in the information society. Revista Economică 67:269–284
23. Helfat CE, Finkelstein S, Mitchell W, Peteraf M, Singh H, Teece D, Winter SG (2007) Dynamic capabilities: understanding strategic change in organizations. Blackwell Publishing, USA
24. Teece DJ, Pisano G, Shuen A (1997) Dynamic capabilities and strategic management. Strateg Manage J 18(7):509–533
25. Wójcik P (2015) Exploring links between dynamic capabilities perspective and resource-based view: a literature overview. Int J Manage Econ 45(1):83–107
26. Hair JF, Hult GTM, Ringle CM, Sarstedt M (2017) A primer on partial least squares structural equation modeling (PLS-SEM) (Second). SAGE Publication, Los Angeles, Washington D.C.
27. Henseler J, Dijkstra TK, Sarstedt M, Ringle CM, Diamantopoulos A, Straub DW, Ketchen DJ Jr, Hair JF, Hult GTM, Calantone RJ (2014) Common beliefs and reality about PLS: comments on Rönkkö and Evermann. Organ Res Methods 17(2):182–209
28. Sarstedt M, Ringle CM, Hair JF (2017) Treating unobserved heterogeneity in PLS-SEM: a multi-method approach. Partial least squares path modelling. Springer, Cham, pp 197–217
29. Cohen S (1988) Psychosocial models of the role of social support in the etiology of physical disease. Health Psychol 7(3):269
30. Sullivan GM, Feinn R (2012) Using effect size—or why the P value is not enough. J Graduate Med Educ 4(3):279
31. Neirotti P, Raguseo E, Paolucci E (2018) How SMEs develop ICT-based capabilities in response to their environment: past evidence and implications for the uptake of the new ICT paradigm. J Enterp Inf Manage 31(1):10–37
32. Tseng S-M, Lee P-S (2014) The effect of knowledge management capability and dynamic capability on organizational performance. J Enterprise Inf Manage 27(2):158–179
33. Chih HL, Shen CH, Kang FC (2008) Corporate social responsibility, investor protection, and earnings management: some international evidence. J Bus Ethics 79(1):179–198
34. Quaye D, Mensah I (2019) Marketing innovation and sustainable competitive advantage of manufacturing SMEs in Ghana. Manage Decis 57(7):1535–1553. https://doi.org/10.1108/MD-08-2017-0784

Four-Port Wideband Circular Polarized MIMO Antenna for Sub-6 GHz Band

Alyaa S. Abdullah, Seham A. Hashem, Wisam S. Al-Dayyeni, Mahmood F. Mosleh, and Ihab Hassoun

Abstract Using a circularly polarized antenna is an attractive method for balancing polarization to transmit and receive the antenna. Therefore, this paper explores a wideband 4-port CP multiple-input multiple-output (MIMO) antenna to be used for 5G wireless technologies in the sub-6 GHz range (3.5–5.9) GHz. Also, the polarization diversity between the neighboring radiators is used to achieve high isolation between antenna components, resulting in a low envelope correlation coefficient (ECC). Other MIMO efficiency metrics are analyzed, such as the proposed structure's Diversity Gain (DG), Mean Effective Gain (MEG), and Channel Capacity Loss (CCL). The finding shows that the suggested design's suitability can be considered a possibility for imminent sub-6 5G MIMO applications.

Keywords MIMO antennas · CP · CST studio · Sub-6 5G · Partial ground plane

A. S. Abdullah (✉) · M. F. Mosleh
Computer Engineering Techniques, Middle Technique University, Baghdad, Iraq
e-mail: bbc0036@mtu.edu.iq

S. A. Hashem
Department of Technical Electronic, Technical Instructors Training Institute, Middle Technical University, Baghdad, Iraq
e-mail: dr.seham.ahmed@mtu.edu.iq

W. S. Al-Dayyeni
Computer Techniques Engineering Department, Dijlah University College, Baghdad, Iraq
e-mail: wisam.talib@duc.edu.iq

I. Hassoun
Faculty of Engineering, City University, Tripoli, Lebanon
e-mail: i.hassoun@cityu.edu.lb

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_72

1 Introduction

For the next generation with a Millimeter-Wave band frequency (mm-wave) bandwidth, the unlimited bandwidth will be adopted to match the Fifth Generation (5G) requirements. By allocating more spectrum to arrive quicker with higher quality video and containing multimedia and facilities, 5G is connected to mm-wave [1]. However, the mm-wave band frequency has easily absorbed plants, air gasses, surrounding buildings, and dust particles, so that the impairment causes are the key challenges in such bands [2]. The MIMO antenna system is used to compensate for the high attenuation and dispersion of 5G band frequencies to address this problem. Antennas with functionality MIMO are also the main specifications for 5G technology literature [3, 4]. On the other hand, the MIMO system suffers from a mutual coupling problem because of limited space, which degrades the overall system's performance.

Some researchers are present more than one solution to address the mutual coupling problem. In this context, the researchers proposed a MIMO antenna using polarization diversity to a mutual coupling reduction of less than −25 dB. A lined patch was implemented between the proposed antenna elements to decrease the mutual coupling effect [5]. Another mutual coupling reduction approach has been introduced less than −45 dB using integrated parasitic strips and inverted ground plane structure [6].

The polarization can be classified into three types: Circular polarization (CP), linear polarization (LP), and Elliptical polarization (EP) [7]. It is possible to consider both circular and linear polarization as exceptional elliptical polarization cases [8, 9].

CP is one of the most critical polarization types to improve the capacity system and mitigate the mutual coupling problem [10, 11]. So, such a technique will be considered in this research. On the other hand, CP was achieving by [12], using an antenna consisting of a D-shaped patch and a modified ground plane, which covered spanning WLAN and WiMAX frequency bands. Also, [13, 14] are proposed a dual-band two-element and four-element CP dielectric resonator-based MIMO antenna, respectively, which achieve a good value of diversity parameters. Also, [15] proposed a four-element MIMO antenna system, each with a dual-polarized port, to enhance bandwidth and mutual coupling with a smaller size and work in a C-band frequency that is a promising 5G smartphone application band.

This research was a prototype of a MIMO antenna system concerned with CP that focused on sub-6 GHz applications for 5G smartphone applications. To achieve high fractional bandwidth, the rectangular slot is built up at the right end of the ground plane, compatible with the requisite facilities for future wireless communication.

2 The Proposed Model Methodology

In this research, a proposed CP MIMO system will be presented to use a sub-6 GHz frequency of 5G system. To achieve better design, we will start with a single CP antenna construction and evaluate its performance. Then the second step included the procedure of MIMO construction with an acceptable geometry. CST Studio Suite 2020 is used for the design of each single and MIMO prototype.

2.1 Single Antenna Design

A single CP antenna part (unit cell) was built with a dimension of 25 × 25 × 1.6 mm^3. The proposed model consists of three layers; the first one is the patch layer, while the second layer represents the separation between patch and ground layers. The substrate layer height is 1.6 mm, made of FR-4 with 4.4 permittivities and 0.02 tangent loss. The third layer represents partial ground consisting of a 50 Ω rectangular microstrip feed line connected via SMA connector to the power source and rectangular microstrip that enhance fractional bandwidth. The specifics of the single antenna's geometry, as shown in Fig. 1, and its design parameters are listed in Table 1.

The design procedure included four steps. Each one focuses on three parameters $S_{11} \leq -10$ dB, broad fractional bandwidth, and Axial Ratio (AR) <3 dB.

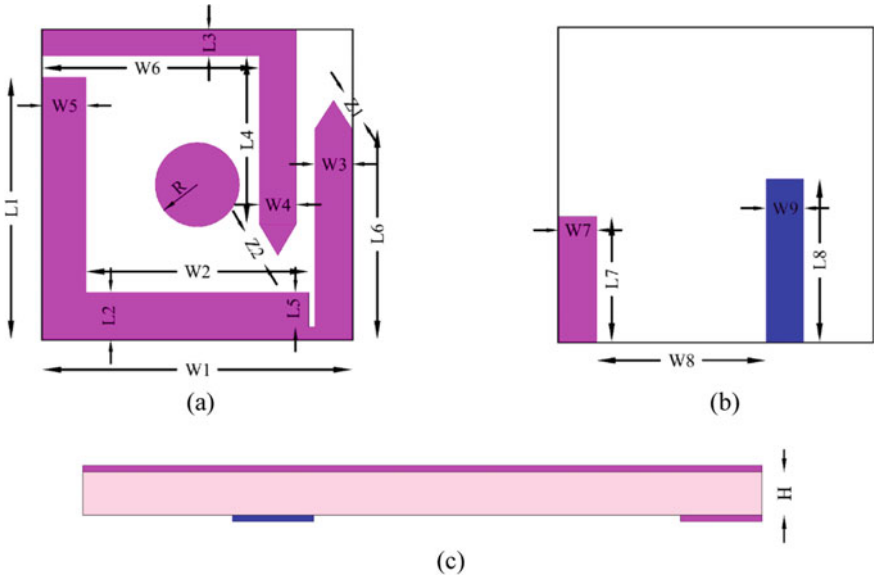


Fig. 1 Proposed unit cell; **a** top view, **b** bottom view, and **c** side view

Table 1 Antenna design parameter values

Parameter	Value (mm)	Parameter	Value (mm)
W1	25	L5	2.8
L1	21.2	W6	17.5
W2	18	L6	17
L2	3.5	W7	3
W3	3	L7	10
L3	2.1	W8	13.5
W4	3	L8	13
L4	13.6	W9	3
W5	3.5	H	1.6
Z1	2.77	Z2	2.94

Step-1: As shown in Fig. 2a, a broad square slot with outer (3.5 mm) is inserted into the square-shaped antenna element's ground plane. A part of the ground's metal level is drilled to form open holes, and two arms facing each other are joined to obtain a CP band.

Step-2: As shown in Fig. 2b, a circle has been added to the design center to enhance low axial-ratio and achieve better circular polarization.

Step-3: Also, two triangles were added to the ends of arms to increase their electrical length to acquire lower frequency bands, as shown in Fig. 2c.

Step-4: Furthermore, Fig. 2d shows the rectangular slot built up at the right end of the ground plane for obtaining high fractional bandwidth and low AR.

2.2 Single Antenna Performance Analysis

After running the simulated single element model, it can be shown from Fig. 3a that the unit cell, which is mentioned in step-1 having resonated with ($S_{11} \leq -10$ dB) for (5.55–6.79 GHz) band. Its polarization is circular when AR = 1.3, which obtain a low fractional bandwidth of 20%.

The results of step-2 show that the low AR is 0.49, as shown in Fig. 3b.

Further, the results of step-3 illustrate that can obtain a lower frequency band of (3.58–6.14 GHz) as shown in Fig. 3a.

Finally, the simulation of the proposed antenna models, as explained in previous steps, shows that the fourth model results in Step-4 represent an optimum model. The resultant of the operating frequency was 5.55 GHz with −25 dB of return losses, the bandwidth of 2.4 (3.5–5.9 GHz), and high fractional bandwidth of 51% at −10 dB. Also, high isolation can achieve the bandwidth of (3–8) GHz, as illustrated in Fig. 4. Depending on [15], the Voltage Standing Wave Ratio (VSWR) value must not exceed 2 dB. According to the VSWR of the designed model approach 1.1 dB illustrated in Fig. 5, the resonance frequency satisfies the VSWR condition.

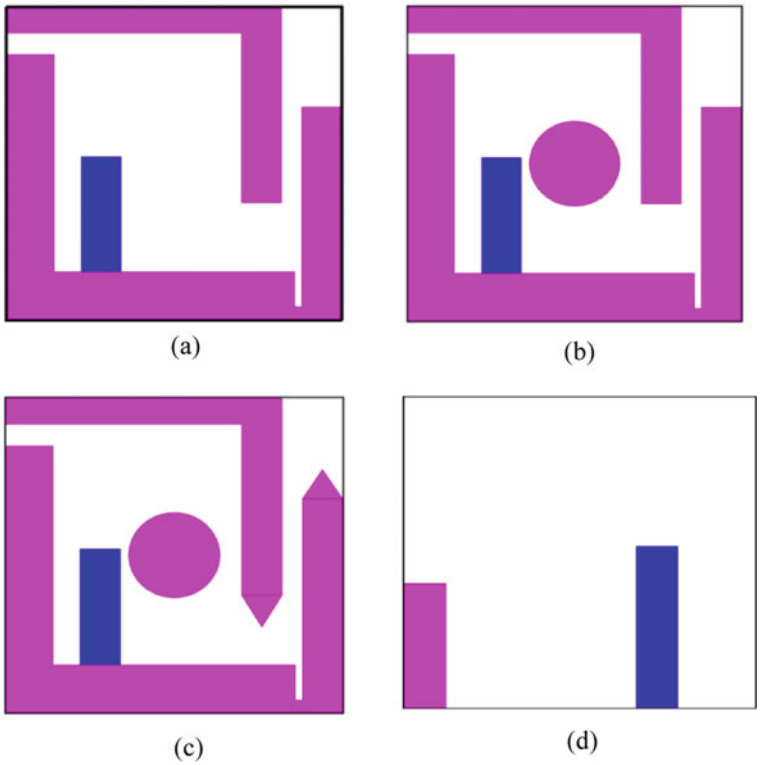


Fig. 2 Evolution steps of the unit cell; **a** Step-1, **b** Step-2, **c** Step-3, and **d** Step-4

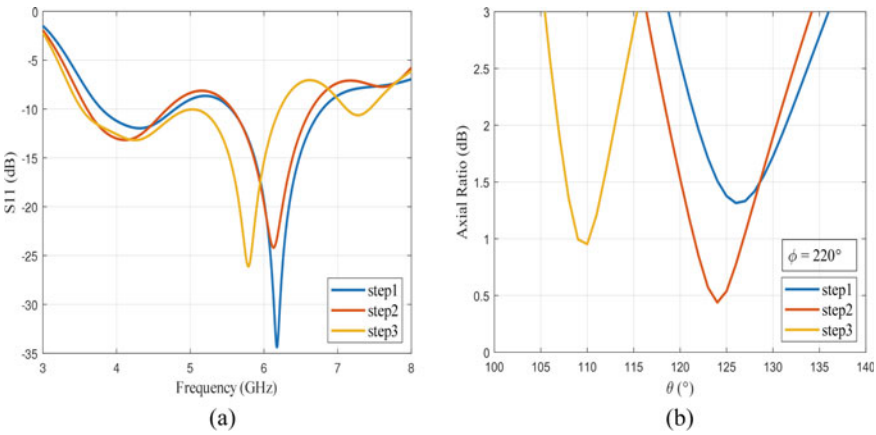


Fig. 3 Simulated S_{11} and AR of the design step; **a** S_{11} of steps 1, 2, 3, and **b** AR of steps 1, 2, 3

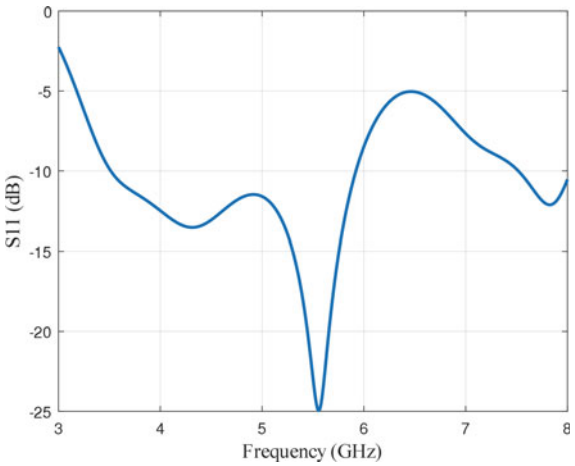


Fig. 4 S_{11} of the proposed unit cell

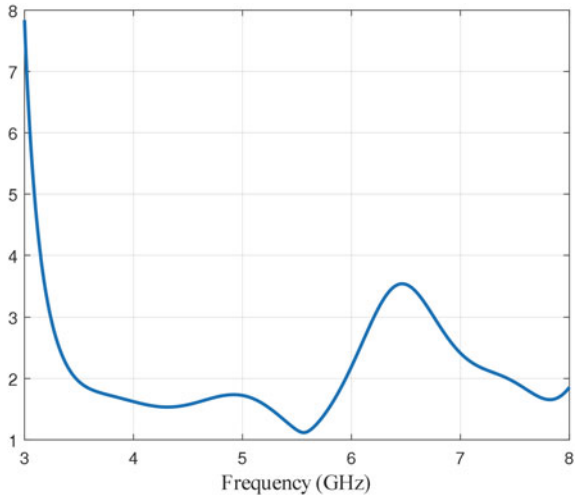


Fig. 5 VSWR of the proposed unit cell

For CP antennas, AR is an essential parameter that helps to quantify the polarization of an antenna "specify the maximum main beam direction with $\emptyset$ and θ values". For antennas have perfect circular polarization AR = 0.37 dB, AR must be less than 3 dB [7], as shown in Fig. 6.

2.3 MIMO Antenna Design

The overall view of the MIMO antenna is delineated in Fig. 7. Four replicas of a single antenna element proposed in (Sect. 2.1) are integrated on a dielectric layer

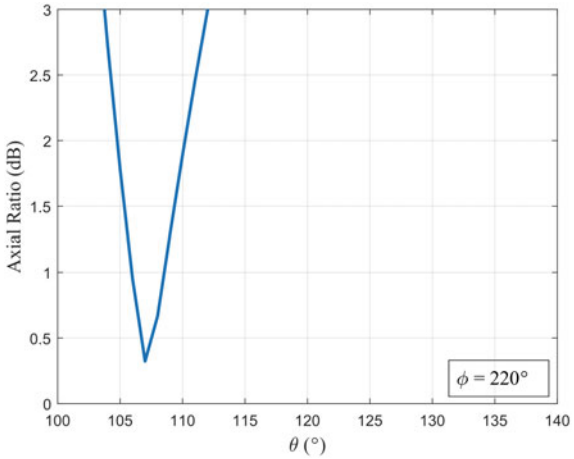


Fig. 6 The 3-dB AR beamwidth of the proposed procedure scheme

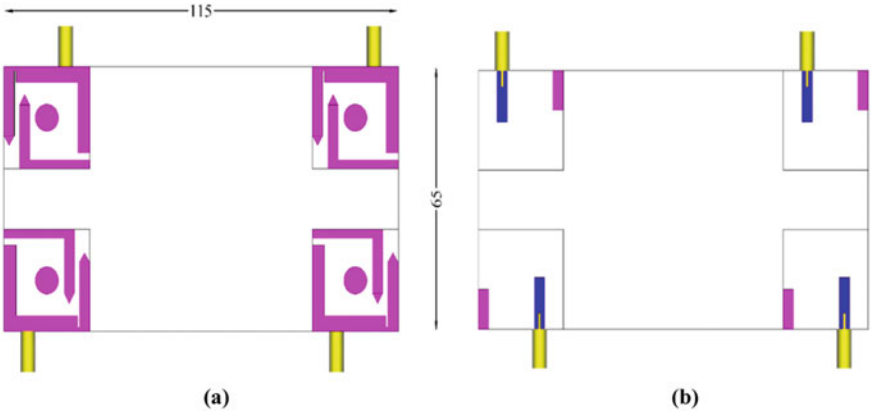


Fig. 7 MIMIO antenna system; **a** top view, and **b** bottom view

Printed Circuit Board (PCB) which made of FR-4 as in every single element substrate layer with a dimension of 115 × 65 × 1.6 mm^3.

2.4 MIMO Antenna Performances and Discussion

For analyzing the MIMO performance, the key parameters that have been studied of the proposed antennas are RL, MC, Envelope Correlation Coefficient (ECC), Diversity Gain (DG), Mean Effective Gain (MEG), and Channel Capacity Loss (CCL). Each parameter will be discussed in the following subsections.

Return Loss and Mutual Coupling

RL is the ratio between power reflected and incident power when $(Sij, i = j)$ as shown in Fig. 8a, which indicates S_{11}, S_{33} = −25, and S_{22}, S_{44} = −30. In contrast, the electromagnetic interaction between the antenna elements in an array is defined as MC $(Sij, i \neq j)$ [16]. The simulated mutual coupling $(Sij, i \neq j)$ between antennas remains below −20 dB for all ports when simultaneously excited, as shown by the results in Fig. 8b.

Envelope Correlation Coefficient

In the case of the MIMO antenna system, ECC is one of the main parameters for calculating diversity, which mainly shows the correlation function between the various ports used simultaneously. Figure 9 shows the ECC curve of the proposed antenna versus frequency. The overall antenna module ensures that the correlation values are lower than 0.5 practical standards, and it is calculated as follows [16].

$$\rho_{eij} = \frac{\left|s_{ii} * s_{ij} + s_{ji} * s_{jj}\right|^2}{\left(1 - |s_{ii}|^2 - s_{ij}^2\right)\left(1 - |s_{ji}|^2 - s_{jj}^2\right)} \tag{1}$$

Diversity Gain

DG describes the impact of the diversity scheme on the radiated power. At the same time, the DG of the proposed MIMO is calculated as a function of frequency by using Eq. (2) according to [8] as shown in Fig. 10. The DG is defined as approximately 10 dB throughout the band, ensuring that the antenna performs well in diversity.

$$DG = 10\sqrt{1 - \left|\rho_{eij}\right|^2} \tag{2}$$

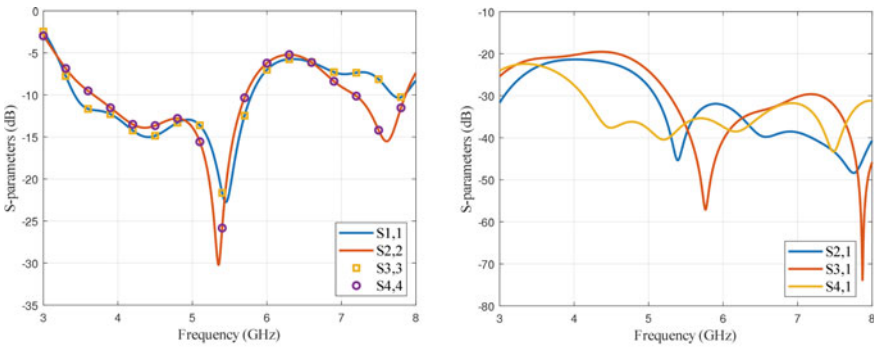


Fig. 8 MIMO antenna system results; **a** return loss $(S_{ij}, i = j)$, and **b** mutual coupling $(S_{ij}, i \neq j)$

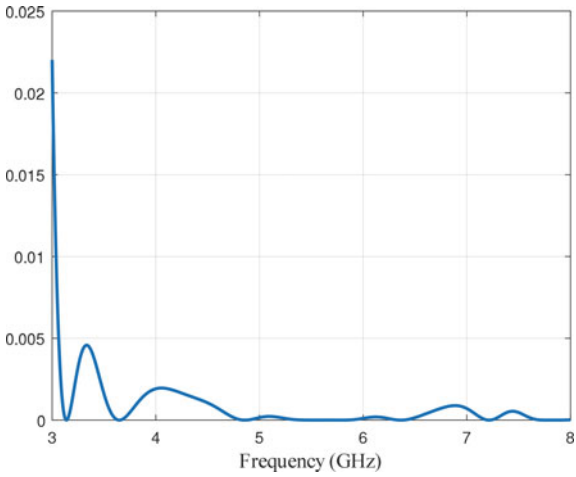


Fig. 9 ECC of the MIMO system

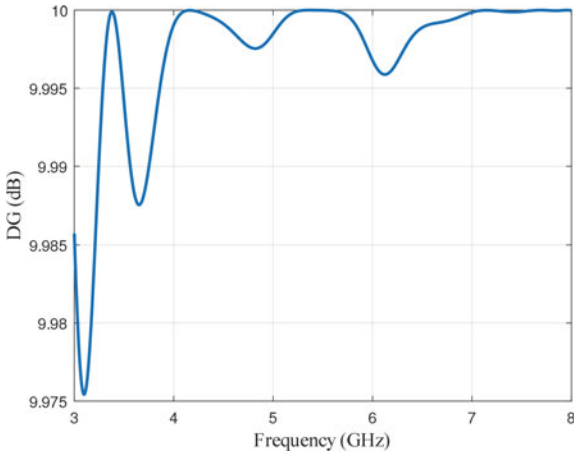


Fig. 10 DG of the MIMO system

Mean Effective Gain

MEG is a significant parameter that means effective gain definition is the main power obtained in fading environments. Figure 11 shows the results of MEG for the proposed MIMO model, which is −3 dB, calculated by [15]:

$$MEG = 0.5\left[1 - \sum_{j=1}^{N} |Sij|^2\right] \tag{3}$$

where N is the number of antenna elements which is equal to 4 in this research.

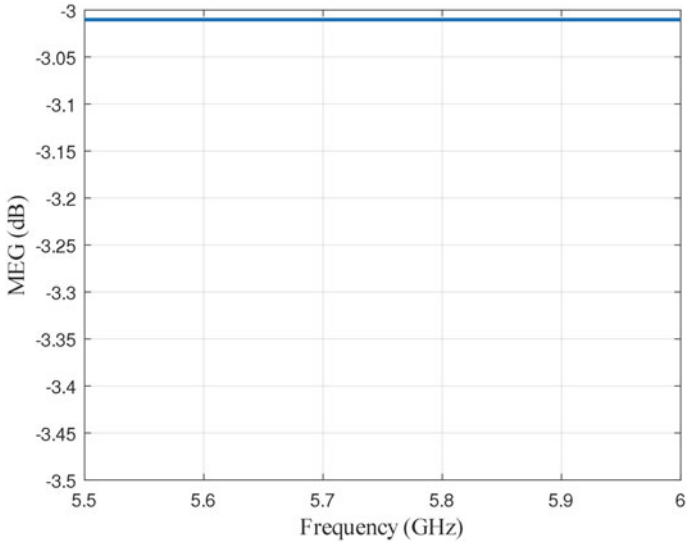


Fig. 11 MEG of the MIMO system

Channel Capacity Loss

Evaluate the MIMO antenna system's performance and provide information on the system's channel capacity losses during the correlation effect. It just depends on the number of antenna elements with a specific assumption. The obtained CCL is less than 0.4 bit/s/Hz practical standards. The CCL calculated numerically by the following equations [17]:

$$C(loss) = -log_2 det(\alpha) \tag{4}$$

where α is the correlation matrix.

$$\alpha = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \tag{5}$$

$$\sigma_{ii} = 1 - \left(|s_{ii}|^2 - |s_{ij}|^2 \right) \tag{6}$$

$$\sigma_{ij} = -\left(s_{ii} * s_{ij} + s_{ji} * s_{jj} \right) \tag{7}$$

Figure 12 depicts the CCL analysis, which observed that it achieves the conditional bound.

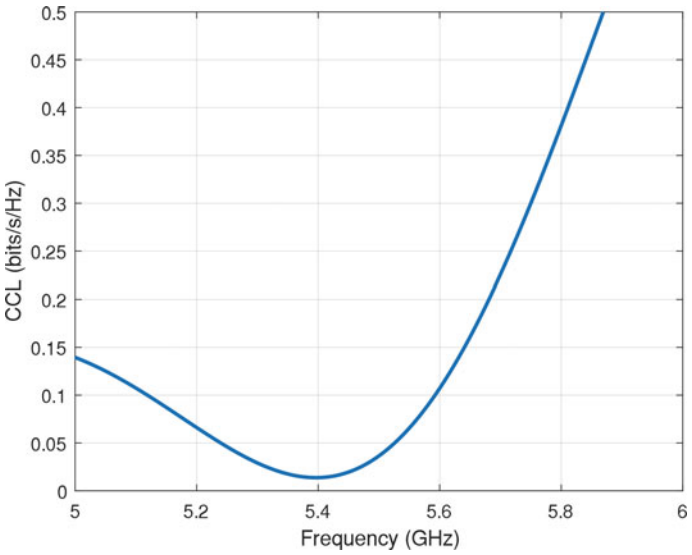


Fig. 12 CCL of the MIMO antenna system

3 Conclusion

This paper proposed a four-port MIMO antenna with CP diversity concentrating on the sub-6 GHz with 5G wireless communication application. The rectangular slot was built up at the right end of the ground plane to achieve high fractional bandwidth 51%, low AR is 0.37, and lacking mutual coupling among ports is lesser than −20 dB. Along with this high DG, also observed low ECC within an operating frequency range. The proposed system's performance prefers to use the MIMO antenna to operate sub-6 GHz with a 5G wireless communication field. Finally, the other results matched the standard requirements mentioned in the considered references.

References

1. Agiwal M, Roy A, Saxena N (2016) Next generation 5G wireless networks: a comprehensive survey. IEEE Commun Surv Tutorials 18(3):1617–1655
2. Talwar S et al (2014) Enabling technologies and architectures for 5G wireless. 2014 IEEE MTT-S international microwave symposium (IMS2014). IEEE, pp 1–4
3. Chakraborty S et al (2020) A 4-element MIMO antenna with orthogonal circular polarization for sub-6 GHz 5G cellular applications (vol 2, pp 1180). Sn Appl Sci 2(8):1–13
4. Zerihun BM, Wondie Y (2017) Massive MIMO for 5G cellular networks: potential benefits and challenges. In: International conference on information and communication technology for develoment for Africa. Springer, Cham, pp 219–227

5. Adam I et al (2019) Mutual coupling reduction of a wideband circularly polarized microstrip MIMO antenna. IEEE Access 7:97838–97845
6. Parbat RS et al (2015) Dual polarized triple band 4 × 4 MIMO antenna with novel mutual coupling reduction approach. In: 2015 IEEE Bombay section symposium (IBSS). IEEE, pp 1–6
7. Saxena S et al (2020) Planar four-port dual circularly-polarized MIMO antenna for Sub-6 GHz band. IEEE Access 8:90779–90791
8. Hussain N et al (2020) Metasurface-based single-layer wideband circularly polarized MIMO antenna for 5G millimeter-wave systems. IEEE Access 8:130293–130304
9. Carrie JS (1999) Development of cavity-backed antennas for use in communications applications
10. Hu C, Dai Y (2016) Design and implementation of a 2.45 GHz circularly polarized microstrip antenna for wireless energy harvesting. In: 2016 5th international conference on environment, materials, chemistry and power electronics. Atlantis Press
11. Yoon W-S et al (2010) A reconfigurable circularly polarized microstrip antenna with a slotted ground plane. IEEE Antennas Wirel Propag Lett 9:1161–1164
12. Kunwar A et al (2019) Circularly polarized D-shaped slot antenna for wireless applications. Int J RF Microwave Computer-Aided Eng 29.1:e21498
13. Bharti G et al (2020) Two-port dual-band circularly polarized dielectric resonator-based MIMO antenna with polarization diversity. Electromagnetics 40(7):463–478
14. Dwivedi AK et al (2020) Design of dual band four port circularly polarized MIMO DRA for WLAN/WiMAX applications. J Electromagn Waves Appl 34(15):1990–2009
15. Al-Ani NMK et al (2020) A Design of MIMO prototype in C-band frequency for future wireless communications. Adv Electromagn 9(1):78–84
16. Kumar P, Urooj S, Alrowais F (2020) Design and implementation of quad-port MIMO antenna with dual-band elimination characteristics for ultra-wideband applications. Appl Sci 10(5):1715
17. Ojaroudi Parchin N et al (2020) Ultra-wideband diversity MIMO antenna system for future mobile handsets. Sensors 20(8):2371

Cryptographic Pseudo-Random Bit Generator Based on New Combination Discrete Chaotic Systems

Muhammad Hamid, Musheer Ahmad, Hussam S. Alhadawi, and Shivam Chandhok

Abstract Traditional 1-D discrete chaotic systems are not suitable to use directly in PRBG design for their cryptographic usage as their structures are simple and have predictability. Pseudo-random sequences have wide applications in image and video encryption, hash functions, spread spectrum communications, etc. In chaos-based cryptography, chaotic systems have been regarded as an important pseudorandom source in the design of pseudo-random bit generators due to its inherent properties of sensitive dependence on initial conditions and parameters. In order to improve the dynamism and features of standard logistic map, a 1-D discrete combination chaos model is proposed in this paper. The chaos model enables to construct new chaotic systems with combination of logistic map and Trigonometric functions. The performance analysis shows that the new systems are more complex and better than the original Logistic map. Further, we also propose to present a new pseudo-random bit generator based on new log-tan chaotic system and log-cot chaotic system. The randomness and other statistic analysis show that our pseudo-random bit generator has good randomness features, satisfy the linear complexity and balancedness requirements well.

Keywords Logistic map · Combination chaotic systems · Pseudo-random number generator · Trigonometric functions · Security

M. Hamid · M. Ahmad
Department of Computer Engineering, Jamia Millia Islamia, New Delhi 110025, India

H. S. Alhadawi (✉)
Department of Computer Techniques Engineering, Dijlah University College, Baghdad 10011, Iraq
e-mail: hussam.alhadawi@duc.edu.iq

S. Chandhok
Department of Electronics and Communication Engineering, Jamia Millia Islamia, New Delhi 110025, India

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_73

1 Introduction

Nowadays security on internet has become an important issue because of the fact that a lot number of users are performing various activities like online bank transactions, e-commerce and e-mail every day. In order to ensure the confidentiality and integrity of the message over the open network like internet, we must produce such a system which should be secure enough, difficult to predict and much efficient to resist various types of cryptanalysis attacks [1]. The cryptographic applications are based on security keys. These keys should be complex enough to predict or to estimate. The cryptographic keys must possess properties like statistical randomness. The randomness property of chaos can be used to generate security keys. The chaotic pseudo random sequence generator algorithm can be a good solution to generate cryptographic keys [2].

In recent decades, chaos has been studied due to nonlinear and difficult to predict phenomenon in nature. Chaos has a great property that it is sensitive dependent on initial condition [3, 4]. One of the most simple and famous chaotic map is logistic map, using this system many cryptographic application has been developed such as algorithm for image encryption and decryption, digital water marking algorithm and pseudo or true random number generator. In the past few years, chaotic map is considered a source of pseudo randomness in the design of pseudorandom bit generators (PRBG) algorithms. In past few years, a huge number of pseudo-random bit generators are design using chaotic maps. It is seen that multi-dimensional chaotic map are difficult to implement as compare to one dimensional chaotic map, hence most favorite chaotic map among all is one-dimensional chaotic maps for chaotic pseudorandom number generator [5]. But traditional 1-D chaotic map is not much secure to use in direct PRBG algorithms because it have a simple structure it can be break easily with phase space reconstruction method [6]. In 1-D chaotic map all the equations are iterative equations in which the next state is depends on the previous state. The morphology of chaotic map is quite simple and can be improved easily by doing some modification in iterative chaotic equations. Here, we design a new chaotic model based on the classical logistic map. Various analysis shows that proposed chaotic systems are much complex than the classical logistic chaotic map, then we present a new PRBG algorithm based on enhanced chaotic systems. The security and statistical analysis shows that our proposed algorithm has good pseudorandom characteristics and highly cable to be utilized for cryptographic applications.

The main contributions of the work done in this paper are as follows.

(1) A chaos combination model is proposed using traditional logistic map and trigonometric functions.
(2) The new 1-D discrete chaotic systems found to have better dynamical features and complexity compared to traditional logistic chaotic map.
(3) Cryptographic pseudo-random bit sequence generator is proposed based on new combination chaotic systems.

(4) The generated binary sequences are tested for randomness through NIST SP800-22 suite. The sequence of one million bits is verified, to demonstrate that the proposed PRBG is highly secure and can provide good statistical randomness characteristics. Generated sequences also satisfy the linear complexity profile and balancedness requirements.

Rest of the paper is arranged as follows: Sect. 2 gives the proposed chaos combination model and its dynamical characteristics evaluation compared to logistic chaotic map. The working of proposed pseudo-random bit sequence generator is discussed in Sect. 3. Whereas, the performance analysis of proposed PRBG is done in Sect. 4 and conclusion of the work is made in Sect. 5.

2 Novel Combination Discrete Chaotic Systems

The mathematical model of well-known classical Logistic map exhibiting the chaotic phenomenon is given by RM May in 1976 [7]. This map has simple structure among all its competitive 1-D chaotic maps available in the literature. It is defined by the equation given as.

$$X_{N+1} = f(u, X_N) = \mu X_N(1 - X_N) \tag{1}$$

The logistic map is dependent on two parameters: the initial value and the bifurcation parameter (X_0, μ). Since, it is an iterative function so the next values are dependent on the values taken from the previous iteration count. By simply varying the parameter μ, the logistic system tends to show the chaos nature when μ lies in [3.57, 4]. Logistic map shows chaotic behavior in between initial values of parameters conditions. $X_0 \in (0, 1)$. But, the logistic map found to possess some demerits including low range of parameter for chaotic behavior, low value of Lyapunov exponents which determines the degree of chaoticity, poor bifurcation of map for parameter, and low entropy content, etc. [8–10]. In order to overcome the limitations, we proposed combination model for new chaotic systems which are defined through of following equations.

$$F_1(X_N, \mu) = X_{N+1} = mod(\mu X_N(1 - X_N) + (1 + X_n)(\tan(\pi X_N)), 1) \tag{2}$$

$$F_2(X_N, \mu) = X_{N+1} = mod(\mu X_N(1 - X_N) + (1 + X_n)(\cot(\pi X_N)), 1) \tag{3}$$

where, $0 < \mu \leq 4$ and $0 < X_N < 1$. This is the combination of logistic map and trigonometric function both are added by multiplied with $(1 + X_n)$ in trigonometric function and the parameter is varied accordingly. The generalized form of new 1-D combination discrete chaotic model involving both logistic and trigonometric functions is shown in Fig. 1. Here, we combined the Logistic map and trigonometric functions such as combining Tangent function with logistic map to get

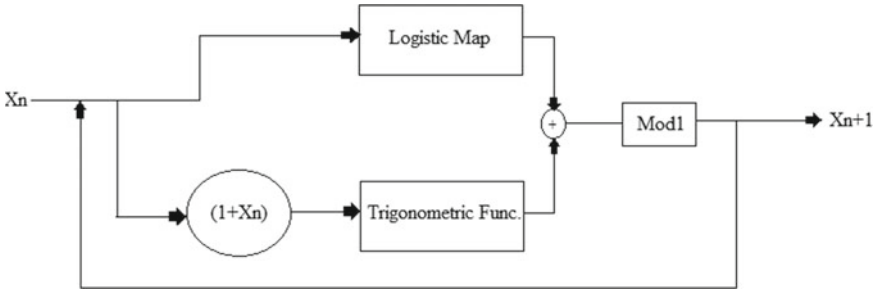


Fig. 1 General model of 1-D discrete combination chaotic system

Logistic-tan system (Eq. 2), and combining Co-Tangent function with logistic map to obtain Logistic-cot system (Eq. 3).

2.1 Lyapunov Exponent

Lyapunov exponent is used to determine chaotic behaviour of a dynamic system. It is a commonly accepted metric which describes separation between two trajectories starting from extremely close initial points. A positive value of Lyapunov exponent indicates chaotic nature, that is, no matter how close the initial distance is, trajectories will diverge exponentially over time making them unpredictable [11, 12]. Hence, a larger value of LE indicates superior chaotic performance of the system. The Lyapunov exponents of mentioned three 1-D discrete systems (Eqs. 1–3) have been determined by varying the system parameter μ. The lyapunov spectrums for three systems are shown in Fig. 2. It is clear that the two chaotic systems start showing chaotic behaviour for $\mu > 2.65$ and has consistent chaotic phenomenon after that unlike logistic map.

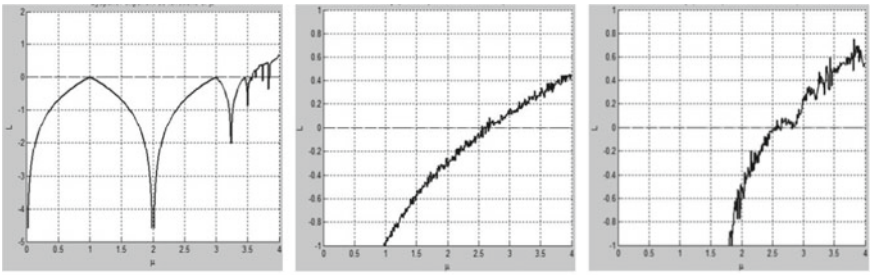


Fig. 2 Lyapunov exponent: **a** logistic map, **b** logistic-tan system, **c** logistic-cot system

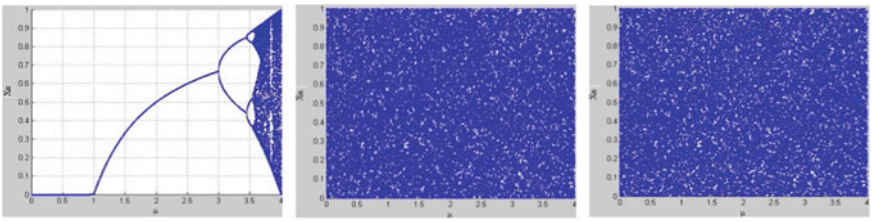


Fig. 3 Bifurcation diagrams: **a** logistic map, **b** logistic-tan system, **c** logistic-cot system

2.2 Bifurcation

Bifurcation analysis is used to quantify regions of chaotic behaviour of a nonlinear dynamic system. Bifurcation is a logical method to analyze behaviour where system dynamism changes with change in parameter and time. Figure 3 shows the bifurcation behaviour of the three chaotic systems [13]. The bifurcation of proposed chaotic systems is more complex, better, and covers the whole area of the plot than logistic map.

2.3 Approximate Entropy

Approximate Entropy (ApEn) is a mathematical algorithm created to measure irregularity in time series. A small ApEn value represents deterministic time series, whereas larger value represents randomness, complexity and unpredictability [14]. The ApEn of time series generated by three chaotic systems for different μ values when $X1(0) = X2(0) = 0.4212$, embedded dimension $m = 2$, time delay for down sampling $\tau = 0.5$ are shown in Fig. 4. The ApEn comparison analysis demonstrates that both the proposed chaotic systems have better complexity, unpredictability and entropy than logistic map.

3 Proposed PRNG Based on New Discrete Chaotic Systems

This section presents the proposed pseudo-random bit sequence generation algorithm which is based on the new 1-D combination chaotic systems. Suppose, the two chaotic systems are denoted as $F_1(X_1, a_1)$ and $F_2(X_2, a_2)$, where $X_1(i)$ and $X_2(i)$ are two initial values (when $i = 0$), a_1, a_2 are the system parameters of $F_1(.)$ and $F_2(.)$, individually. For simplicity, we take $a_1 = a_2 = \mu$ in this paper. The

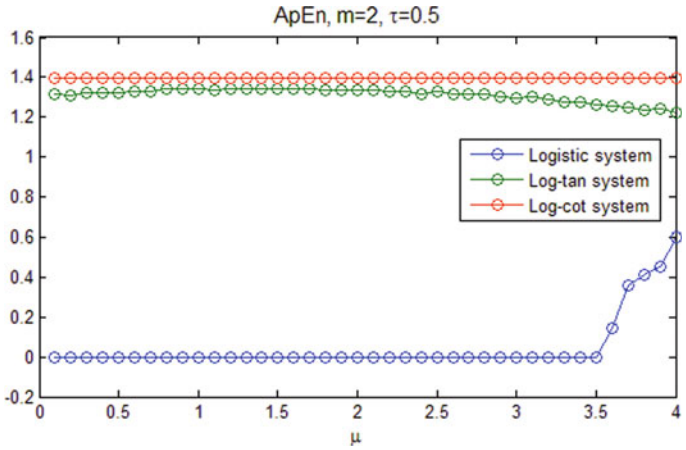


Fig. 4 Comparison of approximate entropies

proposed PRBG has the following operational steps to generate random binary sequences. This algorithm generates 1-byte (8 bit) in each of its iteration.

Step-1: Initialize the seed value of $X_1(i)$, $X_2(i)$, parameter μ and number of iterations.
Step-2: Extract $Q(i)$ from each chaotic variable X_1 and X_2 using Eq. (4) as

$$Q(i) = floor\{(X_1(i) + X_2(i)) \times 10^{15}\} mod(256) \tag{4}$$

Step-3: Covert number $Q(i)$ obtain from step2 in 8-bit binary sequences.
Setp-4: Append the 8-bit block of sequence to the binary sequence.
Setp-4: Repeat *Step*-2 and *Step*-4 for number of iterations to generate PRBS of required length.
Step-5: Declare the final pseudo-random binary sequence.

The proposed PRBG and working is also demonstrated through the block diagram shown in Fig. 5. Here, the gain K shown in diagram is set at 10^{15} (see Eq. 4).

4 Performance Results of Proposed PRNG

There are various statistical test suits to measure the randomness of bit sequence. However, the most widely adopted standard test is defined by the NIST [15] that contains sufficient number of independent statistical tests and detects any deviation

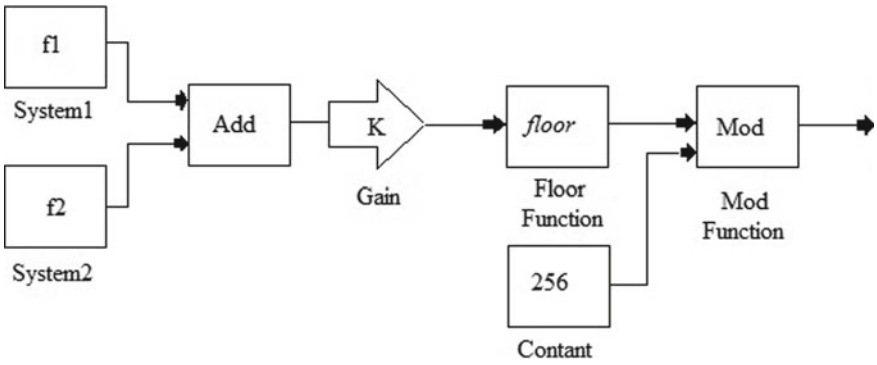


Fig. 5 Working of proposed PRBG using two new combination chaotic systems

from the randomness. The randomness limit for every test of NIST is 0.01, which implies its 99% of all tests finishes test if irregular bits are genuinely arbitrary. The set p-value $\geq$ 0.01 imply that arrangement would be thought to be irregular with a certainty of 0.99. In our experiments, we have created parallel arrangements with length 1000,000 bits generated from proposed pseudo-random bit generator. We test the breezing through proportion of every test in NIST test and consistency of the p-values in every test provided into NIST suite. The outcome is presented in Table 1. A random process means lack of pattern or predictability in events, and it is a sequence of random variables describing a process whose outcomes do not follow a deterministic pattern, but follow an evolution described by probability distributions. The obtained randomness results in Table 1 show that the generated sequences have sufficient randomness as all p-values are adequately higher than 0.01.

Linear complexity of any sequence under examination is the length of shortest linear feedback shift register which can generate the sequence [16]. The Berlekamp Massay algorithm is used to compute the linear complexity. We calculated the linear complexities of binary sequences from proposed PRBG, the obtained values are shown in Fig. 6a. It is clear from the Figure that linear complexity for our PRBG is very close to $n/2$ for sequence length n as expected.

Further, in order to satisfy the Golumb's postulate of balancedness for random sequences, we have also computed the number 1's in the generated binary sequence from our PRBG [17]. The obtained scores are plotted against the sequence length n and shown in Fig. 6b. We can see that the number of 1's is around 50% for different length n. Hence, the proposed PRBG also satisfy the requirement of balancedness as both number of 0's and 1's are equal with very negligible deviation.

Table 1 Randomness test results from NIST SP800-22 suite

Name of test	p-value	Result
Approximate entropy test	0.599052	*Success*
Block frequency test	0.636585	*Success*
Cumulative sum test	0.925027 (forward) 0.898063 (backward)	*Success*
FFT	0.686381	*Success*
Frequency test	0.832107	*Success*
Linear complexity test	0.077597	*Success*
Random excursions test	X = −4 0.348428 X = −3 0.628647 X = −2 0.797521 X = −1 0.622833 X = 1 0.035370 X = 2 0.236948 X = 3 0.409112 X = 4 0.253255	*All success*
Random excursions variant test	X = −9 0.490043 X = −8 0.486339 X = −7 0.525417 X = −6 0.539932 X = −5 0.147636 X = −4 0.068800 X = −3 0.266973 X = −2 0.288028 X = −1 0.284664 X = 1 0.428520 X = 2 0.692597 X = 3 1.000000 X = 4 0.820847 X = 5 0.544331 X = 6 0.304973 X = 7 0.226015 X = 8 0.304119 X = 9 0.678008	*All success*
Longest runs of 1' test	0.942913	*Success*
Overlapping template of all ones test	0.484121	*Success*
Rank test	0.468947	*Success*
Runs test	0.378835	*Success*
Serial test	0.663508 0.400243	*Success*
Universal test	0.244918	*Success*
Non-overlapping templates test	0.996119	*Success*

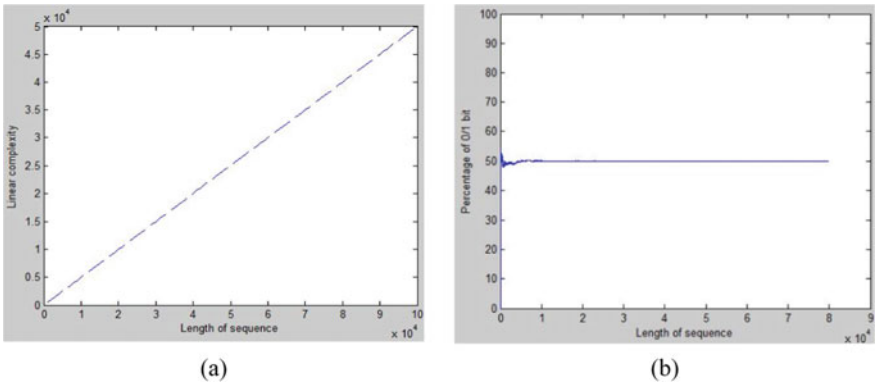


Fig. 6 Statistical analysis of generated sequence: **a** linear complexity, **b** 0/1 balancedness

5 Conclusion

In this paper, we developed a generic model of constructing 1-D discrete chaotic systems to overcome the demerits of established Logistic map. The chaos model is combination of trigonometric function and classical logistic map. The performance analysis demonstrates that our new combinational chaotic systems have better dynamical characteristics than the logistic map. Besides, we also proposed a novel pseudo-random binary sequence generator based on new chaotic systems. The performance insights and security investigations demonstrate that proposed PRBG algorithm consist of good randomness attributes, near to ideal linear complexity and balancedness behaviour. As a future work, the succession of anticipated chaos piece can be utilized for digital image encryption and hash functions designs, and the proposed chaotic system can also be utilized as a part of design of dynamic S-box for AES-like block ciphers.

References

1. Alhadawi HS et al (2020) Globalized firefly algorithm and chaos for designing substitution box. J Inf Secur Appl 55:102671
2. Alhadawi HS et al (2019) Designing a pseudorandom bit generator based on LFSRs and a discrete chaotic map. Cryptologia 43(3):190–211
3. Ahmed HAS et al (2018) Pseudo random bits' generator based on Tent chaotic map and linear feedback shift register. Adv Sci Lett 24(10):7383–7387
4. Ismail SM et al (2018) A new trend of pseudorandom number generator using multiple chaotic systems. Adv Sci Lett 24(10):7401–7406
5. Alhadawi HS et al (2020) A novel method of S-box design based on discrete chaotic maps and cuckoo search algorithm. Multimedia Tools Appl 1–18
6. Short KM (1994) Steps toward unmasking secure communications. Int J Bifurcat Chaos 4 (04):959–977

7. May R (1976) Simple mathematical models with very complicated dynamics. Nature 261:459–467
8. Alzaidi AA et al (2018) A new 1D chaotic map and \beta-hill climbing for generating substitution-boxes. IEEE Access 6:55405–55418
9. Alzaidi AA et al (2018) Sine-cosine optimization-based bijective substitution-boxes construction using enhanced dynamics of chaotic map. Complexity
10. Singla P, Sachdeva P, Ahmad M (2014) A chaotic neural network based cryptographic pseudo-random sequence design. In: 2014 fourth international conference on advanced computing & communication technologies. IEEE
11. Alawida M, Samsudin A, Teh JS (2020) Enhanced digital chaotic maps based on bit reversal with applications in random bit generators. Inf Sci 512:1155–1169
12. Ahmad M, Farooq O (2011) Chaos based PN sequence generator for cryptographic applications. In: 2011 international conference on multimedia, signal processing and communication technologies. IEEE
13. Tutueva AV et al (2020) Adaptive chaotic maps and their application to pseudo-random numbers generation. Chaos Solitons Fractals 133:109615
14. Pincus SM (1991) Approximate entropy as a measure of system complexity. Proc Natl Acad Sci 88(6):2297–2301
15. Rukhin A et al (2001) A statistical test suite for random and pseudorandom number generators for cryptographic applications. Booz-allen and hamilton inc mclean va
16. Murillo-Escobar M et al (2017) A novel pseudorandom number generator based on pseudorandomly enhanced logistic map. Nonlinear Dyn 87(1):407–425
17. Golomb S (1982) Shift register sequences. Park Press, Laguna Hills, CA Aegean

E-Wallet Awareness: The Expansion of Knowledge in Rural Area in Terengganu Malaysia Due to Coronavirus

Nurul Izzah, Mazni Mohamed, Farah Dilaila, Siti Nur Fathiah, and Fify Syazwana Amelda

Abstract The Coronavirus pandemic is a turning point for the payment industry, as those businesses have invested heavily in digital and mobile prosperity. During the coronavirus pandemic, digital payments kept economies going and helped to minimize people's exposure to the virus. E-wallet is an electronic device or online service that enables electronic transactions to be made by a person. The applicant can store funds and that money will be used to make payments to pay taxes, to buy food, to book a ticket, and some others. It is still new to the Malaysian especially at rural areas for instance in Kemaman, Terengganu. Consumers in rural areas are still using cash payment rather than e-wallet due to the lack of awareness and knowledge of the system. This study found the factor affecting the intention to use the e-wallet based on the theory of acceptance and use of technology which are social influence, perceived security, perceived ease of use, and perceived usefulness. **Methodology**: This study used a qualitative and quantitative method such as interview and observation (qualitative) to verify of questionnaires (quantitative). The respondents were from 18 to 46 years and above who use debit card or cashless payment to make transaction. Then for analysing data we used SPSS20. **Findings**: The study found that 89% of respondents know and use e-wallets for shopping, prepaid/postpaid payments, and other utilities. The result showed that the factor of perceived ease of use of e-wallet is high compared to others. This means that the Malaysia rural community also knows that e-wallet application eases their way of doing some purchasing online because they were able to buy anytime without worry anymore about Coronavirus infection. However, e-wallet is still challenging for elderly to understand the process from the beginning of the download of the application, setting it in the smart-phone, choosing the product or shop, and controlling their cash flow. **Application**: Today, people need a fast, simple, and secure way to get everything's done right. Therefore, the easiest and most secure way to buy and pay is through e-wallet. Instead of knowing, people at the rural area are interested in using e-wallet as the improvement bring enjoyable and comfortable

N. Izzah (✉) · M. Mohamed · F. Dilaila · S. N. Fathiah · F. S. Amelda
Department of Business Studies, Faculty of Computer Media and Technology Management, UCTATI, 24000 Kemaman, Terengganu, Malaysia
e-mail: nurulizzah@uctati.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_74

instruction provided. Thus, shoppers also need to prepare all types of payment transactions if there are people who choose to pay with e-wallet instead of cash or card.

Keywords e-wallet · Pandemic · e-commerce · Rural Malaysia · Online shopping

1 Introduction

Malaysia has 13 states, with Terengganu as one of those states. Terengganu is situated the in eastern Peninsular Malaysia, and the east by the South China Sea. Most of the places in Terengganu is a rural area. In a bid to foster digital inclusivity, the Malaysian government attempted to construct and strengthen all projects to connect rural areas to the global e-commerce boom. Supported by e-commerce giants like Shopee and Lazada, the majority of projects enable rural-based entrepreneurs to participate in cross-border supply chains by providing technical support and knowledge in product packaging and e-commerce services [1] including electronic payment such as e-wallet.

Digital wallet is also known as e-wallet refers to an electronic device or online service that allows an individual to make payment through electronic transactions [2]. The applicant will store funds and that money can be used while making payments in order to pay bills, ordering foods, booking a ticket, and many others [3]. It still new to Malaysian as not many people used it differ within the populous city, such as Kuala Lumpur. Due to the lack of awareness, cash is always more convenient than e-wallet. However, people these days are adapting to a new way of life which requires rapid more in every way.

All the privacy information needs to be completed by the user to create a password. This application is recommended because there has a high-security level of it. For example, e-wallet usage help prevents robbery, especially to ladies and elderly. E-wallet stored money in the current account and application. If the mobile phone were robbed, the application would never be opened as it provides a high-security system with a thumbprint and security code [4].

In moving towards cashless country by 2020, there are almost 40 types of e-wallet licensed by Bank Negara Malaysia (BNM), with mobile-based payments being a core to the blueprint. Almost all essentials business such as supermarket, food and beverage businesses, and also small vendors are now opt for e-wallet, starting 2019. It seems like these mobile wallets will become a central part of the nation's payments landscape. For the future implies e-wallets will have a lot of advantages to offer its users.

Then, at the end of 2019, the whole world was shaken by Coronavirus infection [5], and there's no exception for Malaysia. Due to the increased case on the 18 March 2020, Malaysia's new Prime Minister, Tan Sri Muhyiddin Yasin has decided to do the Movement Control Order (MCO) to break the Coronavirus-19 chain. Starting from 14 days as the first MCO, it's dragged on until 3 months. So the

government directed all Malaysian and foreigners have to self-quarantine with a stay at home concept. Due to the coronavirus Coronavirus-19 outbreak and resulting MCO in Malaysia, Malaysia's sectors had downgraded and a lot of citizens lost their works [6].

The Coronavirus-19 pandemic represents a turning point for the payment industry, as those companies have invested heavily in digital and mobile thrive. During the Coronavirus crisis, digital payments have been keeping economies running and helping people to reduce contact with virus. Thus, this situation has increased the promotion and marketing of the use of e-wallet applications in rural areas.

To prevent contagious diseases dispersal, the use of less-contact transaction such as facial-detection, [QR] codes scanning or near-field communication [NFC] is crucial. Virtual transaction offers a more convenient shopping experience for the customer to even do it from their homes and just wait for the products at their doorstep. With initiatives given by the government especially on virtual and e-commerce, it will escalate the potentials of businesses in all areas as well as rural [1]. Some incentives which is given straight to consumers will give even greater help in stimulating buying power, for instance, China gives the citizen vouchers through WeChat Pay that can be redeem immediately [7].

2 Literature Review

2.1 UTAUT

Cashless society is a society where no one uses cash or banknote, where all purchases being made by electronic format [1]. Nowadays, internet access is widespread especially through computer, smartphones and tablet and it becomes necessity product to the society. So this is the right time for launching more electronic forms as they could bring more benefits to society [8]. Thus nowadays, innovative technologies have led to the emergence of a whole new wave of cashless payment technologies. Payment has become even more convenient.

UTAUT stands for unified theory of acceptance and use of technology, that most previous studies are using this model to apply technology acceptance research widely [9]. The UTAUT aims to explain user intentions to use an information system and subsequent usage behavior based-technology. The UTAUT attempts to clarify the user's purpose in using information system and actual actions in use. The theory holds that four main constructs exist: (1) expectation of results, (2) expectation of commitment, (3) social impact and (4) conditions facilitation.

Most researchers review that the intention of use through UTAUT is added with two external variables that are culture and perceived protection (e.g., [3, 9, 10]). Culture would understand user's preferences and perceived protection to illustrate how stable the mobile payment system is via the business market in the area. Users

highlighted these two because it has had a lot of effects on their decision to use the technology [11]. Using the same model, the research influences behavioral intent such as social impact, habit, and circumstances, while price value, the expectation of effort, and hedonic motivation were found to be the least important [10].

Hence, in critical situations such as Coronavirus-19 infection, there must be wisdom behind this disaster [5], where a technology like the UTAUT model promoted should be applied. This really is a perfect time for the government to promote the use of e-wallets as a whole in all areas, especially rural areas. Therefore, Malaysia trying make a step on going to cashless payment, it has many benefits for a nation [3], and Malaysia's government is keen on creating the infrastructure to help its residents and businesses in making that transition [8]. The Government has made some movement to promote the usage of e-payment by provide such as e-PENJANA and e-Tunai Rakyat. Therefore, there are two objectives set for this research:

RO1: To study the awareness level of e-wallet usage in rural area, Kemaman, Terengganu.

RO2: To identify the relationship between factors for intention to use e-wallet system among consumers in rural area, Kemaman Terengganu.

2.2 *Social Influence*

Within the context of rural consumers, social influence mainly refers to influence and inspiration driven by opinions from people around them. Usually, the influence comes from people on the same level of education, salary and has information about e-wallet [4, 8].

Hence, social influence is viewed as the key motivation among rural consumers to adopt new technology, such as e-wallet. A social construct is composed of two basic underlying sets of factors. First are the beliefs that the rural consumer has on their family members or neighbors or friends who they considered as a reference. Second is the motivation of the rural consumer to behave according to the desires of the people of reference [12].

2.3 *Perceived Security*

Perceived security refers to mobile commerce initiatives and further devise appropriate strategies in the planning of payment through electronic. For rural consumers, they are very sensitive to the security level, efficiency, and personal engagement for using the e-wallet [13]. Consumers prefer to have a really safe and secure transaction which is presented in the likert-scale. The e-wallet can store user's money safely as well as they believe that e-wallet companies will protect the user's information [8].

Security is a priority in designing and creating technological system because security it is directly connected with privacy in this modernization system. Privacy and security are also a major concern among the rural consumer for firstly using digital payment. As we know they are new in any technology development, so they are particularly sensitive about these two factors [9].

This study employed privacy and security variable to examine the effects on behavioral changes to the consumer. Privacy and security are significantly one of the prerequisite dimensions that e-wallet providers should emphasize in order to create a positive intention among consumers which means need to strengthen the privacy and security part in the intended use of the e-wallet system [3]. Besides that, the element of trust has a positive impact on the consumer's intention to use mobile payment [2].

2.4 Perceived Ease of Use

Perceived ease of use is described as the extent to which an individual believes that it would be effortless to use a given new technology or system [14]. Thus, the perceived ease of use represents the perception of the customer on how the latest technology or method has arisen to improve the way customers behave and interact. As such, a prior study suggests two common variables in relation to perceived ease of use, which is the "previous experience" of users and the "facilitating condition" of the new technology [8].

Another variable is convenience, it can be considered as the ease and the comfort to use services or products. Supported by previous studies, the convenience of mobility makes online payment as one of favorite mediums [7, 15]. Today, electronic devices such as computers and laptops are also being used as the platform for internet access. Rather than using huge electronic devices, consumers wisely chose a mobile device in making the transaction of payment because it is easy to carry and it doesn't need a big space to keep it. Thus, it shows that convenience has a significant and positive relationship between adoptions of e-wallet, which is e-wallet is more convenient or easy to use, the rate of adopting e-wallet will increase, and vice versa [16]. Not only that, online payment makes small transactions even smother [7].

2.5 Perceived Usefulness

Perceived usefulness determines the extent to which customers assume that using the new product or device or new services will allow them to better perform their jobs or duties [17]. This factor is motivated by major perceived utility variables that are social impact, perceived pleasure, information, and awareness [18, 19].

Specifically, this study emphasized for this factor on benefits that can be provided especially in the pandemic moment, it saves consumer time in making payments, and increasing a new experience through real shopping but paying online to prevent touching as maximum as can.

2.6 Research Framework

Figure 1 presented the research framework for the study. The research framework is developed by integrating the four important factors that influence the intention of using e-wallet UTAUT model. As is shown, social influence, perceived security, ease of use and usefulness are set as the independent variables with rural consumers as set dependent variable. Therefore, four research hypotheses are developed to address RO2.

H1: Social Influence is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia.
H2: Perceived Security is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia.
H3: Perceived Ease of Use is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia.
H4: Perceived Usefulness is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia.

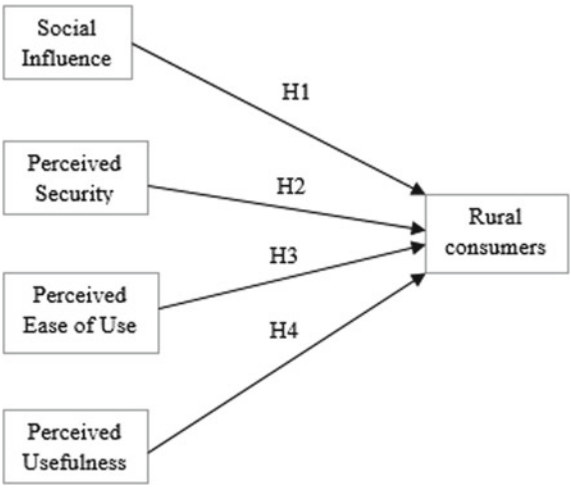


Fig. 1 Research framework

3 Methodology

3.1 Research Instrument

This research used quantitative research approach through a survey. A survey is designed to identify the important factors that influences the intention of using e-wallet and the degree of awareness of intention to use e-wallet among consumers in rural areas in Kemaman, Terengganu Malaysia. The questionnaire consists of 28 questions which is split into Part A, Part B, and Part C as presented in Table 1. In a scale of 1 (strongly disagree) to 5 (strongly agree), the respondents need to rate the dimension of the factors and intention level.

3.2 Population and Sampling

The population for this study is only local people in Kemaman, Terengganu. Based on Department of Statistic Malaysia shown the total number of local people in Kemaman is 110 000. By referring to the result of the study done by Chin [20] and Hair [21], the current study has taken a simple size which is the minimum sample size which was taken by multiplying the number of independent variables by 10. With four (4) variables in the current study, it is resorted into 40 sample size. Taking into consideration the respond rate of previous research done in the rural are in Terengganu is approximately 83%, hence, the targeted sample size for this study is 400 (i.e. 83/100). To avoid demographic bias in sampling and to enhance data validity, 100 respondents from each age of local respondent (i.e. 18–25, 26–35, 36–45, and 46 above) were selected. Thus, the sampling method used in this research is a stratified random sampling method to ensure each segment of the population has an equal chance of being selected [22].

Table 1 Parameters

Section	Dimension	No. of item	References
A	Demographic	4	
B	Awareness level	4	[4, 11]
C	Social influence	5	[4, 11]
	Perceived security	5	[8, 11]
	Perceived ease of use	5	[3, 11]
	Perceived usefulness	5	[8, 11]

3.3 Descriptive Analysis

The normality and reliability of data collected from part B of questionnaire are assessed via Skewness and Kurtosis value and Cronbach Alpha test respectively. Skewness and Kurtosis range of +/−2 represents data is normally distributed. In the meantime, Cronbach Alpha reliability value of greater than 0.60 is suggested to be adequate for testing the reliability of factors [23]. Subsequently, data collected from part B and C of the questionnaire is analysed in steps. Step 1 involved descriptive analysis (mean) to address RO1, while RO2 is addressed via Pearson correlation test at step 2. The evaluation scale for both descriptive analysis and Pearson correction is summarized in Tables 2 and 3.

4 Analysis and Discussions

4.1 Normality and Reliability Test

Normality test in term of Skewness and Kurtosis are conducted on all dependent and independent variables. Skewness and Kurtosis for all variable are within the range of −1.0 to +1.0 suggested that data collected for all variables are normally distributed [23]. A reliability coefficient (Cronbach Alpha) of minimum 0.60 is used as the acceptance level for statistical reliability [21]. Cronbach Alpha for all variables is analyzed and the result shown is above 0.90. Thus, the data collected is statistically significant and can be proceed for the further analysis. The degree of awareness of Intention to Use E-Wallet to address RO1 (To study the awareness level of e-wallet usage in rural area, Terengganu.), the mean score level of the five important factors is ranked and summarized in Table 4.

Table 2 Awareness level rating scale

Mean (μ)	Awareness level
$1.0 \leq \mu < 1.5$	Not very important
$1.5 \leq \mu < 2.5$	Not important
$2.5 \leq \mu < 3.5$	Neutral
$3.5 \leq \mu < 4.5$	Important
$4.5 \leq \mu \leq 5.0$	Very important

Table 3 Correlation rating scale

Coefficient of correlation (r)	Correlation
$0.00 \leq r < 0.20$	Very weak
$0.20 \leq r < 0.40$	Weak
$0.40 \leq r < 0.60$	Moderate
$0.60 \leq r < 0.80$	Strong
$0.80 \leq r \leq 1.0$	Very strong

Table 4 Awareness level

No.	Elements of awareness about e-wallet	Mean (before pandemic)	Mean (during pandemic)
1	Degree of information and knowledge about e-wallet	3.124*	3.942**
2	Perceived enjoyment with e-wallet	2.335^{o}	3.453*
3	Facilitating condition using e-wallet such as prepaid/ postpaid, utilities payment and parking fee	2.518*	3.751**
4	Previous experience using e-wallet	1.356^{oo}	3.335*

^{oo}Not very important, ^{o}Not important, ***Very important, **Important, *Neutral

As referred to Table 4, findings from descriptive analysis revealed that respondents among rural areas perceived the two elements of e-wallet awareness in Kemaman, Terengganu Malaysia during pandemic is important with the mean score range from 3.7 to 3.9. The important level is referred to the degree of knowledge and information about e-wallet, and facilitating conditions during using the e-wallet. This finding is a novelty to show that rural areas in Malaysia have good awareness and knowledge about the development of technology devices. Most consumers are aware of e-wallet because of a critical situation of pandemic Coronavirus-19. However, the result is totally different as compared to before pandemic where, the knowledge about e-wallet and facilitating provided is less. Therefore, this study highlighted that the pandemic condition was also able to influence the rural consumer to find out more about e-wallets. Besides that, based on the good experience they will recommend an e-wallet to someone he/she knows.

Whereas during pandemic the level of perceived enjoyment and experience using an e-wallet is increased with a mean range of 3.3 to 3.4 compared to before. That means, people in rural areas a perceived enjoyment with e-wallet but there are still a lack of stores that provided e-wallet systems. Based on this result it is also proven that almost rural areas accepted e-wallet for their shopping routine especially during pandemic COVID-19. Just a small group in rural areas are still refused to use e-wallet because of a few things such as not ready yet to use a new application, has no motivation to learn new knowledge, and still feel comfortable with the old way.

4.2 Significant Between Factors of Intention to Use E-Wallet Among Rural Consumer at Kemaman Terengganu, Malaysia

To address the RO2 of this study (i.e. To identify the relationship between factors for intention to use e-wallet system among consumers rural area, Kemaman Terengganu.), as well as to test the four research hypotheses, data collected from part B and C of questionnaire is further analyzed via Pearson correlation test.

The purpose of Pearson correlation test is to assess the relationship between the awareness level perceived by respondents on the factors that influences intention to use e-wallet and the perceived level of intention to use e-wallet. The result of analysis is summarized in Table 5.

Based on Table 5, before pandemic COVID-19 there were two hypotheses (H1 and H4) which are not significant, however during pandemic COVID-19 the significant value (i.e. p-value) for all hypotheses are less than 0.05. Hence, this suggested that at confidence level of 95%, the relationship between the factors of UTAUT that influences the intention of using e-wallet among rural consumer in Kemaman, Terengganu Malaysia are significant.

The coefficient of correlation (r) for the four hypotheses are ranged from 0.504 to 0.743, whereby the important factors for Perceived Ease of Use and Perceived Usefulness present a strong correlation with intention to use e-wallet among rural consumers in Kemaman, Terengganu Malaysia. While factors for Perceived Security and Social Influence showed a moderate relationship with the intention to use e-wallet among rural consumers in Kemaman, Terengganu Malaysia.

Hence, from this finding is suggested that the adoption of e-wallet among rural consumers in Kemaman, Terengganu Malaysia is strongly related to items such as easy to understand and to learn about the e-wallet process. This factor is high because they are excited about new applications. Additionally, as announced on

Table 5 Hypotheses result

Hypotheses	Coefficient of correlation (r)		P-Value		Result	
	Before	After	Before	After	Before	After
H1: social influence is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia	0.489	0.504	0.070	0.020	Not Sig.	Sig.
H2: perceived security is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia	0.603	0.617	0.010	0.000	Sig.	Sig.
H3: perceived ease of use is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia	0.628	0.743	0.000	0.000	Sig.	Sig.
H4: perceived usefulness is significant with the intention of using e-wallet among rural consumers in Kemaman, Terengganu Malaysia	0.402	0.718	0.100	0.000	Not Sig.	Sig.

Before: before pandemic COVID-19, After: during pandemic COVID-19

tabling Budget 2020 by the Malaysian government, they have distributed RM 30 incentive for all Malaysian whose eligible, the program which is called E-Tunai Rakyat. E-Tunai Rakyat is made of purpose in promoting the use of digital payment with testing Malaysian perception towards the new technology system in Malaysia. The main purpose of the federal government's initiative to encourage the public, small businesses, and real stores to accept and use digital payments. Therefore, as an introduction to the e-commerce system for people in rural areas, this is the right time to try and learn new technology with a simple method.

Meanwhile, previous experiences of the people in rural areas are also related to the perceived security, as well as the level to which the local people believe that the technical infrastructure is established to support the use of e-wallet technology. Additionally, under social influence factor there list is a four reasons why e-wallet are chosen by the consumer in a rural area moreover in the city also; (i) ease of sending and receiving money, (ii) no need cash in hand, (iii) safe from the robber, (iv) no need to keep a coin for small goods, and (v) absolutely easy compare to the other financial service providers such as banks.

5 Conclusion

In January 2020, before pandemic COVID-19, the Malaysian government introduces the RM30 e-wallet incentive. This program aimed of using e-wallet among Malaysian also to encourage the adoption of cashless payment in Malaysia. However, it received less response especially at rural areas in Malaysia compared to urban places. The factors related mainly to knowledge and education background, had no reasons to use it, and low awareness of e-wallet apps adoption. In fact, this apps are really useful. All the payment and transactions can be made with three types of e-wallet operators which have been partnered with the government, they are GrabPay, Boost and Touch 'n Go E-Wallet. The applicant can choose from one three platforms to apply for their e-wallet or in the Malay language called E-Tunai.

With RM 450 million, following the rules Malaysians aged 18 and above with an annual income of less than RM100, 000 are qualified for the initiative. Other than that, the e-wallet operators would also improve the RM30 free money given by the government, by giving their own incentives through their own benefit. This means that eligible Malaysians will receive free money of more than RM30 through this initiative.

The government is expecting up to 15 million Malaysians would benefit from this government initiative. Khazanah Nasional Berhad (Sovereign Wealth Fund Investment Company belongs to Malaysia) has been appointed to implement the E-Tunai Rakyat as a digital stimulus meant to encourage the use of e-wallets and digital payments and cut back cash circulation, especially during an infection pandemic of Coronavirus-19 each of human activity such as gathered in one place, touching, or exchange goods from others should be reduced as possible as can.

Also, it was an effort by the Malaysia government as a wider push to digitize the economy.

Therefore, this study highlighted four UTAUT factors due to benefits that can be offered particularly in the pandemic moment, such as not wasting time to make payments, and growing a new experience by shopping in actual, but paying online to avoid touching as much as possible.

Besides that, this study revealed the importance of perceived security influencing consumer behavioral intention to adopt a mobile payment system. As a recommendation, this study stressed to e-wallet providers to give more priority to mobile commerce initiatives. Furthermore, devise appropriate strategies in planning to increase consumption of e-wallet especially among rural areas in Malaysia.

Instead of knowing, people are interested in using the e-wallet as the improvement bring enjoyable and comfortable instruction provided. The shoppers also need to prepare all types of payment transactions if there are people who choose to pay with e-wallet instead of cash or card.

References

1. Izzah N, Rifai D, Yao L (2016) Relationship-courier partner logistics and E-commerce enterprises in Malaysia: a review. Indian J Sci Technol 9(9)
2. Wong WH, Mo WY (2019) A study of consumer intention of mobile payment in Hong Kong, based on perceived risk, perceived trust, perceived security and technological acceptance model. J Adv Manage Sci 7(2):33–38
3. Karim MW, Haque A, Ulfy AM, Hossain AM, Anis ZM (2020) Factors influencing the use of e-wallet as a payment method among Malaysian young adults. J Int Bus Manage 3(2):1–2
4. Tan OK, Abdul Aziz F, Ong Choon H, Goh Chin F, Lim Kim Y, Mohd Saiful I, Choi Sang L (2020) E-wallet acceptance among undergraduates in Malaysia. Test Engineering & Management. The Mattingley Publishing Co., Inc., pp 12990–12998. ISSN: 0193-4120
5. Noor Fzlinda F, Pazim Khairul H, Langgat J (2020) The impact of Covid-19 pandemic crisis on micro-enterprises: entrepreneurs' perspective on business continuity and recovery strategy. J Econ Bus 3(2):837–844
6. Papadopoulos T, Baltasb KN, Baltac ME (2020) The use of digital technologies by small and medium enterprises during COVID-19: implications for theory and practice. Int J Inf Manage. https://doi.org/10.1016/j.ijinfomgt.2020.102192
7. Wu F-S, Yen Y-S (2014) Factors influencing the use of mobile financial services: evidence from Taiwan. Scientific Research; Mod Econ 5:1221–1228
8. Teoh Teng Tenk M, Hoo Chin Y, Lee Teck H (2020) E-wallet adoption: a case in Malaysia. Int J Res Commer Manage Stud 2(2):216–233
9. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425–478
10. Mugambe P (2017) UTAUT model in explaining the adoption of mobile money usage by MSMEs' customers in Uganda. Adv Econ Bus 5(3):129–136
11. Junadi S (2015) A model of factors influencing consumer's intention to use e-payment system in Indonesia. Procedia Comput Sci 59:214–220
12. Crespo AH, del Bosque IR (2005) Influence of corporate social responsibility on loyalty and valuation of services. J Bus Ethics 61(4):369–385
13. Silviana Novita Wahyudi A (2019) Analysis of factors motivating in the use of E-Wallets in. Int J Bus Manage Invention (IJBMI) 8(12):56–59

14. Rathore HS (2016) Adoption of digital wallet by consumers. BVIMSR's J Manage Res 8(1):69–79
15. Liu GS, Pham TT (2016) A study of factors affecting the intention to use mobile payment service in Vietnam (David Publishing). Econ World 4(6):249–273
16. Cheryn YX, Kong SY, Lee VA, Ong CP, Lim SY (2018) Moving into cashless society: factors affecting adoption of e-wallet. A Research Project of Bachelor of Business Administration (Hons) Banking and Finance, Universiti Tunku Abdul Rahman, Faculty of Business and Finance, Department of Finance
17. Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 13(3):319–340
18. Nysveen H, Pedersen PE, Thorbjørnsen H (2005) Intentions to use mobile services: antecedents and crossservice comparisons. J Acad Mark Sci 33(3):330–346
19. Shin DH (2009) Towards an understanding of the consumer acceptance of mobile wallet. Comput Hum Behav 25(6):1343–1354
20. Chin WW (1998) The partial least squares approach to structural equation modelling. In: Marcoulides GA (eds) Modern methods for business research. Lawrence Erlbaum Associates, Mahwah, New Jersey
21. Hair JF, Ringle CM, Sarstedt M (2011) PLS-SEM: indeed a silver bullet. J Mark Theory Pract 19(2):139–152
22. Sekaran U, Bougie R (2013) Research methods for business: a skill building approach, 6th edn. Wiley, Chichester, UK
23. Hair JF, Hult GTM, Ringle CM, Sarstedt M (2017) A primer on partial least squares structural equation modeling, 2nd edn. Sage, Thousand Oaks

MySimB': An Interactive Learning Malay Idioms with Augmented Reality

Nur Sukinah Aziz, Azliza Yacob, Saadiah A. Rahman, Nur Liyana Zakaria, Noor Suhana Sulaiman, and Wan Nur Idayu Tun Mohd Hassan

Abstract Currently, learning idioms is included in Level One syllabus for primary school students. However, more often that not students need to traditionally refer to its meaning via physical dictionary. Certain students find learning through this method is outdated and difficult; thus, may tend to demotivate them to learn. Considering the aforementioned issues, an interactive learning tool called, MySimB' was developed using Augmented Reality approach. The main aim for developing MySimB' is to assist primary school students in learning standard Malay idioms easily and effectively. Learning idioms through MySimB' also provides better understanding for primary school students, teachers and parents due to its interactive feature which was attractively designed and developed. With its Augmented Reality approach, students are expected to be more involved with their learning process and have high interest and motivation in learning Malay idioms; thereby, improve students' achievement.

Keywords Augmented reality · Interactive learning · Website · Idioms · Multimedia

1 Introduction

The use of technology such as information communication technology (ICT) helps teachers and students as teaching and learning aids [1, 2]. This is in line with ICT policy for national education which was previously announced by Minister of Education; whereby, ICT has become the main teaching medium, while teachers are its facilitators commencing in 2010. Various technologies have been integrated with numerous educational disciplines. One of them is Augmented Reality (AR), which

N. S. Aziz (✉) · A. Yacob · S. A. Rahman · N. L. Zakaria · N. S. Sulaiman · W. N. I. T. M. Hassan
Faculty of Computer, Media and Technology Management, University College TATI, Kemaman, Terengganu, Malaysia
e-mail: nursukinah@uctati.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_75

allows a combination of real and virtual world. AR is one of the latest potential technologies that has seen diverse applications in the field of education [3–5]. However, compared to other technologies such as multimedia and online learning, studies on the use of AR in education are still in the early stage [6]. Therefore, it is vital that AR and its implementation in education are explored. The use of learning application with high-quality design is crucial and helpful in promoting intrinsic motivation and improving learning quality [7].

Bahasa Melayu or Malay is a mandatory subject for primary and secondary students in Malaysia. Meanwhile, Malay idiom or *simpulan bahasa* is introduced in the syllabus as early as at Standard 3 for primary school students. Generally, by reading books, most students should learn idioms. In fact, websites and applications for idioms are also made available. However, in terms of interface, most websites and applications are less appealing and non-interactive. Interactive modules are not included in such websites and even applications do not apply AR to the idioms. Several problems have been identified when teachers and students utilize current method in teaching and learning idioms. Among those include the need to browse through books such as Dictionary of Proverbs or *Kamus Peribahasa* if they need to search for idioms and its meanings and the need to carry around the dictionary which is burdensome considering the students already have to bring textbooks to school daily. Apart from that, lack of interactive website for primary school students is also considered an issue. Most of these websites are not user-friendly and its interface does not attract primary school students [8]. Therefore, it is imperative that AR components are integrated into learning. This study has developed MySimB' as an educational tool to ease parents and teachers in assisting and teaching students in a fun and interactive learning way in learning Malay idioms.

This paper is organized as follow: Sect. 1 introduces the introduction of the study and Sect. 2 presents related studies pertaining to augmented reality and augmented reality in education. Whilst, Sect. 3 discusses about research methodology; Sect. 4 elaborates on the result and discussion of the study. Finally, Sect. 5 concludes the study.

2 Related Studies

2.1 Augmented Reality

Currently, AR is one of the prevalent technology trends that is only getting bigger as AR-ready smartphones and other devices are becoming more accessible around the world. It is a technology that involves overlaying computer graphics on real world. AR refers to the extension of user's real environment as the content is virtually and synthetically created. This content utilizes various media, including but not limited to text, audio, images and 3D objects [9]. AR is a real-time direct or indirect view of a physical real-world environment enhanced/*augmented* by adding

virtual computer-generated information. It lets people see real-life environment right in front of them with overlaid digital augmentation. Hence, it can be described as a technology that involves overlaying computer graphics on real world and adopting real-time information in form of text, graphics, audio and other virtual enhancements integrated with real-world objects.

Augmented Reality aims at simplifying user's life by bringing virtual information to his immediate surroundings and any indirect view of real-world environment such as live-video streaming. AR enhances user's perception of and interaction with real world [1]. AR defined as implementation by three characteristics as it is: (a) a combination of real world and virtual elements, (b) interactive in real-time, and (c) registered in 3D (i.e., the display of virtual objects or information is intrinsically tied to real-world loci and orientation) [10, 11]. AR also described as systems those which combine "real and computer-generated information in a real environment, are interactive and in real time, and [which align] virtual objects with physical ones" [12]. AR identified as a "human–computer-interaction, which adds virtual objects to real senses that are provided by a video camera in real time" [13].

A fundamental characteristic of AR systems is that it allows the user to see a combined view of virtual imagery and real objects. AR system consists of three simple steps: Recognition, Tracking, and Mix [14]. In addition, any image, object, face, body or space is recognized on which virtual object is superimposed [14]. While tracking real-time localization in the space of image, object, face, body or space is performed, media in a form of video, 3D, 2D, text etcetera is superimposed. Marker-based AR System uses physical-world symbol as a reference point to overlay computer graphics. In the system, camera continuously snapshots a target object and processes the image to estimate the position, orientation and movement of visualization display with respect to the target object. For example, a 2-dimensional printed marker is placed in front of a webcam. Then, computer interprets this symbol to overlay an on-screen graphic as if it is directly on top of the marker in physical world. Hence, AR is able to improve and enhance our perception of the surroundings by combining sense, computer and display technologies.

2.2 Augmented Reality in Education

AR is a new technology with immense capabilities and promising educational significance that enables new approaches to education [15]. AR design that considers the usability of interactive media and its application in developing a highly interactive and structured classroom presentation system [15]. They combine AR with other interactive media as they believe that AR system provides unique learning experience and is a powerful teaching and learning tool.

AR has positive impacts on learners compared to non-AR initiatives by enhancing students' understanding and motivation [16]. High users' enthusiasm with AR engagement shows greater satisfaction, more fun and a desire to revisit the experience. Others advantages of AR; whereby, improved learning curve and

increased motivation are more than 20% compared to the aforementioned benefits [17]. In selected studies by [18], most of AR applications in education field include those of natural sciences, mathematics and statistics (52.50%). social sciences, journalism and information (15%), arts and humanities (15%), and engineering, manufacturing and construction (15%).

3 Research Methodology

This study employed system development in the process of which ADDIE methodology was implemented. ADDIE model is a common method that instructional designers and training developers commonly used to develop a system [19]. "ADDIE" stands for **A**nalyze, **D**esign, **D**evelop, **I**mplement, and **E**valuate. The methodology consists of five (5) phases: Analysis, Design, Development, Implementation and Evaluation. The analysis phase clarifies the problems and goals of instruction. It also identifies learners' learning environment, existing knowledge and abilities. Important information required in the application development is gathered and recorded as document of requirement specification. At the design phase, MySimB' and AR for MySimB' were designed for students to learn Simpulan Bahasa via simple instructional guide and exercise. The idioms used at the website were previously referred to *Kamus Peribahasa*. Meanwhile, interface design for the website first version was revised because it did not include AR as the module. Figure 1 displays the first version of MySimB'.

Fig. 1 Interface of MySimB' main page first version

The interface needs to be restructured to include AR as the main module. The second version of MySimB' has four (4) main modules which include *Simpulan Bahasa*, *Carian*, *Uji Minda* and AR. The remaining modules are list of idioms with its meaning and sentence examples called *Simpulan Bahasa* module and searching module called *Carian*. Prior to this, two (2) types of learning ga mes module in the first version were developed. Nevertheless, the second version of MySimB' combined several learning games into one (1) module. This is according to [20–22], the component of gamification can be used as a strategy that able to increase the level of students motivation and participation. Gaming can be a strategic design choice for maximizing learner attention and knowledge retention, and subsequently minimize their boredom and distraction [23]. Therefore, MySimB' website still maintains its interface of four (4) main modules. This phase is critical because new interface for the second version of MySimB' with AR needs to be developed. The interface design is crucial because it needs to focus on multimedia elements inside the AR.

In development phase, three (3) main development processes occur including interface design, website and AR. Website interface and image for AR were designed using Adobe Illustrator CC. While it might be tempting to use fancy layouts and different design patterns, it simply makes it crowded and busy looking for the users. Simple interface allows users to see clearly the goods or services offered on the website particularly for children. It definitely looks phenomenal to develop a complex design interface; however, it seems likely to cloud the underlying intent. The best way to improve user interface design is to stick to simplicity [24]. The website was developed by employing Adobe Dreamweaver CC with HTML 5, JavaScript, Cascading Style Sheet (CSS) and PHP. Whereas, the AR effects were developed using Unity software. Figure 2 displays welcome interface for MySimB'.

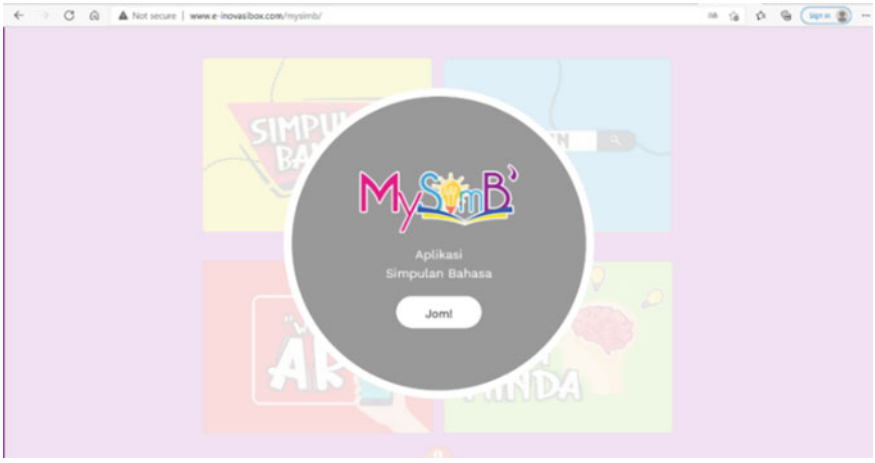


Fig. 2 Welcome interface of MySimB'

Figure 3 shows the main interface of MySimB' second version that has four (4) modules. The modules are *Simpulan Bahasa* that has a list of idioms and their meanings. Over 500 idioms arranged alphabetically (A to Z) were added along with their meanings by referring to *Kamus Peribahasa* published by Oxford Fajar. This is to ensure the contents are ideal and can be referred by local and international users especially primary school students. Whilst *Carian* module or searching module is for users to search for idioms; *Uji Minda* module is for them to test their knowledge via two (2) types of learning games. Meanwhile, "with" AR module consists of ten (10) images for AR. As a result, users can search for specific idioms easily and interactively via games-based learning concepts.

Figure 4 exhibits "with" AR module interface. It has ten (10) idioms whereby users can select any idiom.

When users click on the intended idiom, an image box soon appears as shown in Fig. 5. The box displays a combination of images for each word in the idiom.

When users scan the image, the AR appears as indicated in Fig. 6. Users can click on the sound icon and sound related to the idiom's explanation will be heard. In this AR, users employ visual and listening skills to enhance their learning process. It engages users in learning idioms in more interactive and interesting method. Learning material and tool in MySimB' play a supporting and simplifying role in the teaching–learning process which include facilitating effective relationship between student and vision lesson and continuing to keep the student interested in the subject [25].

MySimB' was launched recently and users can use the website via this link http://www.e-inovasibox.com/mysimb/index.html. The application for MySimB' AR can be downloaded from Google Play Store as shown in Fig. 7. At the moment, this application can only be used by android users.

Fig. 3 Interface of MySimB' second version main page

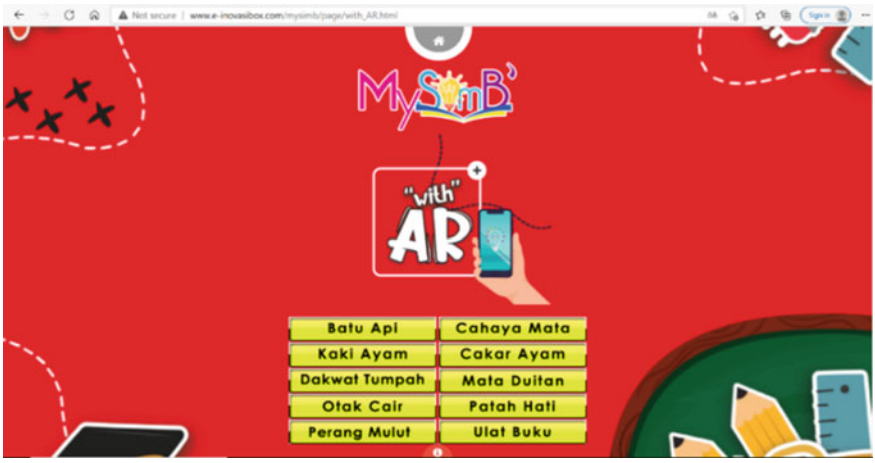


Fig. 4 AR module interface

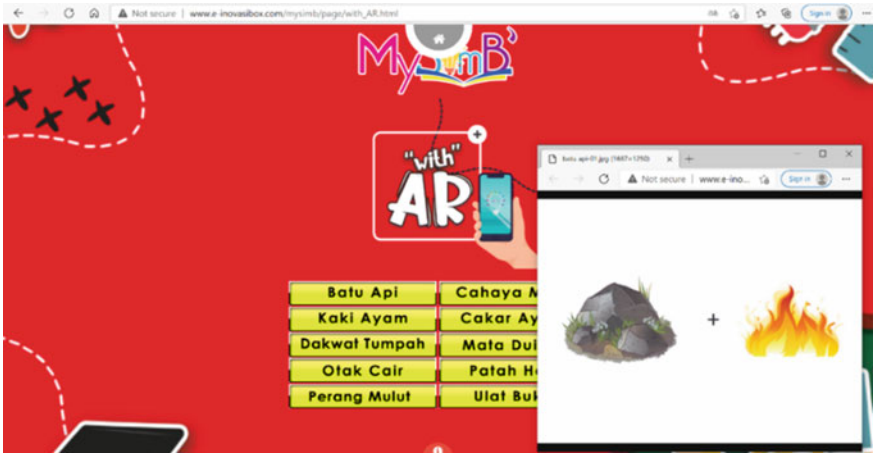


Fig. 5 Interface after users click on the button of intended idiom

Users' evaluation of MySimB' was expected to be completed by primary school students, teachers and parents after MySimB' testing was administered in several schools in Terengganu, Malaysia. Those schools were Sekolah Kebangsaan Gong Balai, Marang, Terengganu and Sekolah Kebangsaan Kg. Baru Kerteh, Kemaman, Terengganu. However, due to COVID-19 concerns, the evaluation is temporarily unavailable. Nevertheless, the researchers will conduct pre-test and survey evaluation with the aforementioned stakeholders after Malaysian government lifts Recovery Movement Control Order (RMCO) in Malaysia.

Fig. 6 An example of AR that appears when users scan the image

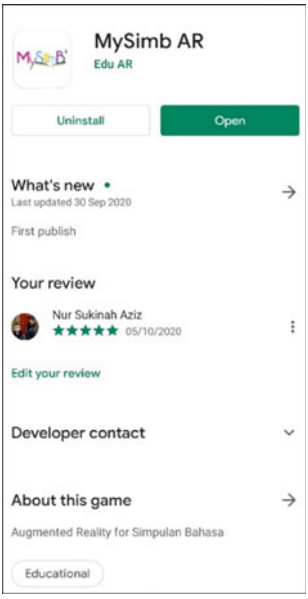


Fig. 7 MySimB' AR Application can be downloaded from Google Play Store

4 Discussion

In this study, MySimB' with the use of Augmented Reality approach is developed for conducting standard *Malay* idioms learning activities. Learning idioms becomes more attractive and effective when students are provided with multimedia material that enables to enhance students' learning interest. Apart from engaging students in learning process, MySimB' also can increase students' motivation towards learning standard *Malay* idioms by including gaming components inside the learning tool.

It has been proven by many researchers that AR and gaming components developed for educational purposes are able to increase students' interest and motivation to learn. Among main design principles of engaging people to learn are: interactive and interaction of the games provided, adaptive feedback, have a clear winning criterion and reward [26–28]. Comparisons between existing idiom learning materials were made. One of the conclusions that can be drawn is that there are still gaps that need to be filled in improving primary school students' interest for learning idioms. Although, many learning resources related to idioms are available; yet, it is insufficient. Learning idioms should be presented in a more attractive and interactive way. With MySimB', users can learn idioms either through a laptop, computer or mobile phone which supports responsive design. Furthermore, MySimB' has been designed to provide simple instructional guidance and exercise for students to learn idioms.

5 Conclusion

Augmented Reality (AR) has been widely utilized in many educational contexts to enable students to learn in engaging, fascinating, and effective It also has been applied to various areas such as marketing, entertainments, mathematics and statistics, social science, tourism, surgery, mathematics and statistics, social sciences, journalism and information, arts and humanities, engineering, construction, logistics, manufacturing and maintenance. The adoption of AR in learning is driven by having easy and affordable access to powerful modern devices (e.g., mobile phones) that can be used as an efficient delivery medium for learning. AR technology also has shown positive impacts on primary school students' motivation. Hence, this proves that AR environment is able to help them to develop better understanding on learning content. MySimB' is an effort to enhance idiom's understanding among students, which can be utilized as an educational tool to ease parents and teachers in assisting and teaching students in a fun and interactive learning way. Thus, it is hoped that MySimB' is capable to engage students in their learning process, increase their motivation towards learning idiom and eventually improve their achievement.

References

1. Masood T, Egger J (2020) Adopting augmented reality in the age of industrial digitalisation. Comput Ind 115:103112
2. Alalwan N, Cheng L, Al-Samarraie H, Yousef R, Ibrahim Alzahrani A, Sarsam SM (2020) Challenges and prospects of virtual reality and augmented reality utilization among primary school teachers: a developing country perspective. Stud Educ Eval 66:100876
3. Cai S, Liu C, Wang T, Liu E, Liang JC (2020) Effects of learning physics using augmented reality on students' self-efficacy and conceptions of learning. Br J Educ Technol 0(0):1–17

4. Kravtsov H, Pulinets A (2020) Interactive augmented reality technologies for model visualization in the school textbook. CEUR Workshop Proc 2732:918–933
5. Estudante A, Dietrich N (2020) Using augmented reality to stimulate students and diffuse escape game activities to larger audiences. J Chem Educ 97(5):1368–1374
6. Ibáñez MB, Uriarte Portillo A, Zatarain Cabada R, Barrón ML (2020) Impact of augmented reality technology on academic achievement and motivation of students from public and private Mexican schools. A case study in a middle-school geometry course. Comput Educ 145
7. Cao X, Wang F, Zheng Z (2012) The experimental research on elearning instructional design model based on cognitive flexibility theory. Phys Procedia 25:997–1005
8. Aziz NS, Sulaiman NS, Zakaria NL, Yacob A, Hassan WNITM, Rahman SA (2020) MySimB': interactive tool for learning Simpulan Bahasa through educational website. In: The 1st international recent trends in engineering, advanced computing and technology conference (RETREAT) 2020
9. Carmigniani J, Furht B (2011) Augmented reality: an overview. In: Handbook of augmented reality. Springer, pp 3–46
10. Azuma RT (1997) A survey of augmented reality. Presence Teleoperators Virtual Environ 6(4):355–385
11. Zhou F, Dun HBL, Billinghurst M (2008) Trends in augmented reality tracking, interaction and display: a review of ten years of ISMAR. In: 7th IEEE international symposium on mixed and augmented reality 2008, ISMAR 2008, pp 193–202
12. Höllerer TH, Feiner SK (2004) Mobile augmented reality. In: Telegeoinformatics: location-based computing and services, Taylor & Francis Books, pp 200–221
13. Yuen SC-Y, Yaoyuneyong G, Johnson E (2011) Augmented reality: an overview and five directions for AR in education. J Educ Technol Dev Exch 4(1)
14. Amin D, Govilkar S (2015) Comparative study of augmented reality SDK's. Int J Comput Sci Appl 5(1):11–26
15. Dalim CSC, Kolivand H, Kadhim H, Sunar MS, Billinghurst M (2017) Factors influencing the acceptance of augmented reality in education: a review of the literature. J Comput Sci 13(11):581–589
16. Radu I (2012) Why should my students use AR? A comparative review of the educational impacts of augmented-reality. In: ISMAR 2012—11th IEEE international symposium on mixed and augmented reality 2012, science and technology papers, pp 313–314
17. Diegmann P, Schmidt-Kraepelin M, Van den Eynden S, Basten D (2015) Benefits of augmented reality in educational environments—a systematic literature review. In: 12th international conference on Wirtschaftsinformatik, pp 1542–1556
18. Garzón J, Pavón J, Baldiris S (2017) Augmented reality applications for education: five directions for future research. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) LNCS 10324:402–414
19. McGriff SJ (2000) Instructional system design (ISD): using the ADDIE model. Instr Syst Des Using ADDIE Model 68–68
20. Tillmann N, De Halleux J, Xie T (2013) Teaching and learning programming and software engineering via interactive gaming (2013 35th)
21. Hamzah WMAFW, Ali NH, Mohd Saman MY, Yusoff MH, Yacob A (2015) Enhancement of the ARCS model for gamification of learning. In: Proceedings—2014 3rd international conference user science engineering experience engineer engage, i-USEr 2014, September, pp 287–291
22. Wan Hamzah WMAF, Haji Ali N, Mohd Saman MY, Yusoff MH, Yacob A (2015) Influence of gamification on students' motivation in using e-learning applications based on the motivational design model. Int J Emerg Technol Learn 10(2):30
23. Giannakas F, Papasalouros A, Kambourakis G, Gritzalis S (2019) A comprehensive cybersecurity learning platform for elementary education. Inf Secur J 28(3):81–106
24. Kurilovas E, Kubilinskiene S (2020) Lithuanian case study on evaluating suitability, acceptance and use of IT tools by students—an example of applying technology enhanced learning research methods in Higher education. Comput Human Behav 107

25. Aziz NS, Yacob A, Sulaiman NS, Rahman WNITHSA, Zakaria NL (2020) Development of MySimB': a motivational aspect. In: The 1st international recent trends in engineering, advanced computing and technology conference (RETREAT) 2020
26. Sahin D, Yilmaz RM (2019) The effect of augmented reality technology on middle school students' achievements and attitudes towards science education. Comput Educ
27. Chang KE, Zhang J, Huang YS, Liu TC, Sung YT (2020) Applying augmented reality in physical education on motor skills learning. Interact Learn Environ 28(6):685–697
28. Cai S, Liu E, Shen Y, Liu C, Li S, Shen Y (2020) Probability learning in mathematics using augmented reality: impact on student's learning gains and attitudes. Interact Learn Environ 28(5):560–573

A Review of Cyber Security Awareness (CSA) Among Young Generation: Issue and Countermeasure

Noor Suhana Sulaiman, Azliza Yacob, Nur Sukinah Aziz, Nooraida Samsudin, Wan Ainul Alyani Wan Mohamed, Saadiah A. Rahman, Wan Nur Idayu Tun Mohd Hassan, Akhyari Nasir, Syahrul Fahmy Abdul Wahab, and Wan Roslina Wan Othman

Abstract Cybersecurity has emerged as an essential concept in everyday life, requiring the involvement of individuals. Although cybersecurity is empowered by government to all level of user, as a critical demanding situation confronted, however, the visibility and public focus stays limited especially to young user. The engagement of Internet is regularly taken into necessity for sharing information, learning, transactions and controlling the physical world, mainly during pandemic Covid-19. Hence, Cybersecurity Awareness (CSA) is a key defense in the protection of user and cyberspace. This review paper elaborates the CSA issue and methodology that had been done by other researcher, consequently the encounter effort taken to enhance the CSA among user, focus on young generation.

Keywords Cybersecurity awareness · Online learning · Young generation · Information security · Internet engagement

1 Introduction

Over the last decade, a growing amount of attention has been paid to investigating aspects of insider threat. Every day, as technology progresses, our world becomes more connected than it has ever been. Although these developments make our daily lives simpler, they also put our personal information at risk. When most people make an online purchase, check their email, or go to the store, they do not worry about their identities being stolen [1]. People and organizations have become increasingly vulnerable to cyber-attacks, which have had a significant negative

N. S. Sulaiman (✉) · A. Yacob · N. S. Aziz · N. Samsudin · W. A. A. W. Mohamed · S. A. Rahman · W. N. I. T. M. Hassan · A. Nasir · S. F. A. Wahab · W. R. W. Othman
Faculty of Computer, Media and Technology Management, University College TATI, Kemaman, Terengganu, Malaysia
e-mail: suhana@uctati.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_76

effect [2]. Furthermore, the high Internet penetration rate and low security awareness among users make it an enticing destination for cyber criminals. The advent of the Internet, as well as the use of numerous online applications and exposure to ever-evolving social media, has exposed students to online danger. Online fraud, cyberbullying, and phishing are only a few of the dangers that students face on a regular basis [3]. Modern day society is driven by technology. While advantageous, online technology is not without its challenges, and one of these is the emergence of cybercrime. Nowadays, children use the Internet every day for various purposes, including playing online games, social media or even for online learning, especially when the Movement Control Order (MCO) that was issued by the Malaysian central government as a measure to help stop the Covid-19 network. The role of the Internet is very important for most children without realizing the risks online, also what action taken properly to online risk. Insider risk can be caused by two types of behaviour: Intentional and Unintentional behavior [4]. Internet is now as a necessity requirement to every student to attend online learning. Consequently, the impacts are all amount of information have been shared online for both socialization and education, also all of this is possible, and it makes life easier and more fun, but without control over the computers, the workplace's infrastructure is vulnerable to cyberattacks [5]. Thus, cyber Security is important to educate all user especially young generation especially when they using Internet. Cyber security is refers to a measure for protecting computer systems, networks, and information from disruption or unauthorized access, use, disclosure, modification or destruction [6]. Another definition is as the protection of cyberspace itself, of the tangible or intangible technologies that support cyberspace, of electronic information, and of the users in their personal, societal and national capacities [7].

Meanwhile, with the often Internet usage, the number of cases of information privacy violation by young Internet user occurs year after year. According to reports, the most common age for cyber bullying is 13–15 years of age [8]. Three out of four young people say they have been bullied in online, and 89% of them got no action taken [9]. Some of them had uncomfortable images of them, they were shared online and some were victims of hate speech. Not only are these young people bullied online [10], but It also faces cases of identity theft from infringement of data privacy. Around a million kids were in 2017 as young users are easily a target for their poor judgment when it comes to identity fraud and scammers. Data privacy awareness is still new, so it is important to concentrate on it. The rapid technological shifts are the reasons why knowledge is relevant are the greatest use of the Internet among young generation. There are many potential risks and threats posed by Internet harassment and inappropriate usage of personal use of information [11].

Thus, this paper aim is to and discuss the CSA issues and countermeasure taken by government to spread and nurture the cyber security awareness in coping the information privacy among young generation. High engangement of Internet usage nowadays is a motivation to responsible ourselves nurturing the user especially young generation in necessity of cyber security awareness. This paper is divided into 5 sections. Section 1 for introduction; Sect. 2 on literature review; Sect. 3 deals

with cyber security awareness among young generation: case study review; Sect. 4 discusses the effort on cyber security awareness for young generation, and lastly Sect. 5 on conclusion.

2 Literature Review

Modern day society is driven by technology. While advantageous, online technology is not without its challenges, and one of these is the emergence of cybercrime. Individuals who plan to target young children emotionally or financially are likely to come across them if they are unprepared to defend themselves from online threats. Bullied children may be more vulnerable to fraud because they seek friendship online and are exploited. In combating child identity fraud, guardians must proactively control their child's finances by constantly tracking conduct, even if they believe nothing has happened; reviewing statements online; and using account notifications, especially those provided to mobile devices.

Meanwhile, the Covid-19 pandemic has forced global shutdown of several activities, including educational activities [12]. The COVID-19 pandemic has created considerable uncertainty, anxiety, and a drastic change as regards our way of life [13]. As a result, there has been a massive crisis-response migration of colleges, high schools, and primary schools to online learning as the educational portal via Google Meet, Zoom, Whatsapp, education learning website and many more. Hence, digital device and Internet is a necessity to digital learning produce more engagement of Internet to young generation in learning and other Internet activities, including playing games, communicate in social media and etc. Many youngsters are compelled to be online, not only socialization through social media but for educational purposes as well.

During the period of the Movement Control Order (MCO), the Prime Minister has recommended online learning initiatives at home. For the first phase of the MCO, a large number of students from all corners had difficulty in starting online learning sessions. This is due to several factors that are considered to be among causes of online learning failure during pandemic. It is important to enhance parental control of their children's use of social media. The monitoring process should begin at home, so parents must be attentive and constructive. If parent not be able to track their children's online activities at home, the children would be exposed cyber threats such as pedophiles, fraud, data breach and cyber bully [14]. As an example, Internet addiction is a major problem due to interferes with a person's development of self-identity. According to the World Internet Statistics Review, almost 78.8% significant number of Internet users in the country, especially children are suffer from severe Internet addiction [15]. According to the survey, only 4 out of 10 parents are aware of their children's online activities. Parents are often not aware to children's access and this consequence the exposure to inappropriate web sites, lead them to cyberspace's dangers.

Consequently, virus which threatens human life comes at an unexpected time, even without any preparation. Malware is the most common tool used in cyberspace to carry out malicious intents, either by exploiting existing vulnerabilities or by using specific characteristics of emerging technologies [16]. The government's announcement to implement teaching and learning at home is inevitable and this may lead to the emergence of various issues including technical issues. To enable the smooth online learning, students need a smartphone or computer with fast Internet access. However, not all students have it, in fact some have to borrow or take turns with a large number of siblings. Internet access issues are also sometimes less encouraging disrupting the learning process.

2.1 Information Privacy

As the information privacy issues becomes the highlight in recent news, the access to the existing data are widely available. The findings show that the number of cases of Internet scams, online abuse, cross-site scripting, and identity theft has risen as a result of a lack of information about the fundamentals of Internet knowledge. Malaysia has recorded 10,000 cases of cyber bullying in 2018 and the cases increasing day by day. A comprehensive and relevant legal framework should be strengthened time to time in combating the issues [10]. In previous, the important of information privacy was explained by Warren and Brandies [17]. Their understanding was the consequences impacted by interfering other's confidential information and publish the information to others. According to Black's Law Dictionary, information privacy defines as a right of an individual to determine to what extend his information is communicated to others [18].

2.2 Risks Faced by Young Generation in Unprotected Information Privacy

Establish statistics, recent news and headlines indicate a range of violations of information among the young generation. In Malaysia by 2020, there are 1,780,662 of IP address that infected with malicious software, also Digital identity is 100% more common to young Internet users under 25 than it is to older Internet users 55 and above [19] as in Fig. 1.

In year 2016, the first Parental Intelligence System in the world has encountered a lack of 6.8 million private text databases messages, 1.8 million photos and 1,700 children's profiles in depth [20]. The information was available for at least 48 days until the service provider shut it down. Then, in 2017 German government banned the Internet-connected or smart toys named "My Friend Cayla" doll [21]. This is due to the concerns over hackers that it could be used to spy on children. The privacy of children has been comprised and sadly it comes from a toy. Meanwhile,

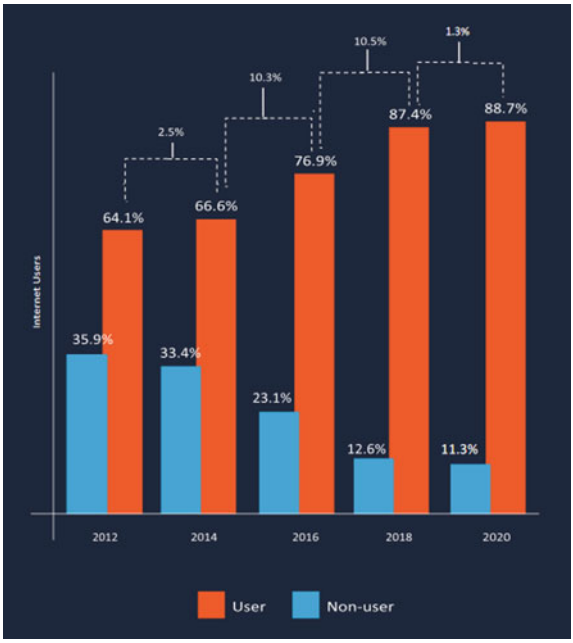


Fig. 1 Internet user's across the years

Google's YouTube has to pay $170 million fine in 2019 because the site failing to obtain parental consent in collecting children's data under the age of 13 [22]. According to Federal Trade Commission (FTC), the organization has violated the privacy laws of children. It was abused by the safety of these children from the unauthorized collection of data. In these cases, it illustrates how important it is to educate young generation in protecting themselves.

3 Cyber Security Awareness Among Young Generation: Case Study Review

In research of Rahim et al. [23], Kirkpatrick's Four Learning Evaluation Model was used as a conceptual structure for their assessment, which consists of four phases of sequential assessment: Phase 1—Reaction, Phase 2—Learning, Phase 3—Conduct, and Phase 4—Result. The evaluation was conducted using a mixed method testing approach, and the instruments used in Phases 1 through 4 were a survey, a pre-test, and a questionnaire. The results of this study showed that young people have a positive reaction to the programme content, and that they have reported improvements in their awareness and skills, as well as the practice of desired actions regarding personal data security. However, undesired behaviour was also observed among children, suggesting that the current module used to educate children needs to be changed. The improvements include:

(i) Decision making process in using personal data,
(ii) Management of online application,
(iii) Management of online contents,
(iv) Management of password and username.

In order to provide insight into the efficiency level of the current module used to educate young people and help they decide on a better module, the outcome of this evaluation is important for stakeholders.

In Jun et al. [24], the study looked at whether online children can understand online risks and what the best ways are to mitigate the risks to which they might be exposed. Researchers discovered that the cyber security awareness programme is important and helpful in educating children about online safety. In Maria et al. [25], a well-prepared security awareness programme is important. It should be specific and actionable. To be approved, a cybersecurity awareness programme should be adaptable by different cultures.

Rahman et al. [26] report highlighting research studies carried out in the cybersecurity sector of education. Several Databases (Emerald, Google Scholar, EBSCOhost, Sci, and Scopus) keywords such as cybersecurity, data protection, and cyber schooling have been discussed. The chosen literature was contained in the two languages that the researcher would understand, Malay and English, respectively. It was found based on a synthesis of the selected literature, it is critical to protect children through cybersecurity education so that they are aware of the possible dangers they may face when using Internet contact tools such as social media, online gaming, and chat. Cybersecurity education, however, has many obstacles. This includes teacher knowledge, as well as a lack of skills, support, and resources. All concerned stakeholders, including educators, parents, peers, and the government, must work together to determine the best strategy for protecting children from cybercrime and cyberbullying through school-based cybersecurity. Cybercrime against children and adolescents is certainly a concern for parents, as they sometimes do not realise their child is a victim of cybercrime.

Muniandy et al. [27] stated that heavy Internet users are students in higher education, and they will form the workforce of the future. Learners can be educated on cyber protection for accidents using such variables. Although increasing cyber security incidents cannot be prevented by education and training, Internet users must be trained to raise their awareness of these events so that they can respond appropriately to take precautionary measures.

Jambari et al. [28] use a quantitative approach that included a survey and an online questionnaire. There were 872 parents of children under the age of 17 who took part in the survey. The following questions will be addressed by this research:

(i) How well-informed are parents about home cyber threat?
(ii) What is extent of parent preparation for their children to employ BYOD in structured learning?

The findings indicate that parental understanding of cyber threats is moderate, and that it needs to be increased in order to encourage cyber safety. Parents can

develop a greater understanding of cyber security if they are exposed to cyber threats at a young age.

4 Effort on Cyber Security Awareness for Young Generation

Various studies have been performed in the past, and several cyber security based gaming apps have been developed in recent years to educate users, especially children; for example, "Cybersecurity Lab," which is intended to teach basic cyber security skills to young people; "Internet Protection," a web-based Internet safety game; and "Kids game—FBI," which is about online security management. However, there is one thing that the group needs to discuss. Many games and applications, for example, are often part of research projects and are created routinely or quickly, however after their review, they often vanish and are seldom accessible to the general public [29]. Thus, [30] employing Design-Based Studies in investigating cyber security related learning experiences of children in informal settings (DBR). The goal of this research is to improve existing teaching techniques and methods as well as to discover new creative and innovative teaching methods, particularly in the CSA field.

Bada et al. [31] state that there has not been the desired effect attempts to enhance security policy and foster a healthy society. Therefore, it is crucial to focus objectively on constrain to improve the behaviour of residents, customers and employees about information security. Changing behaviour necessitates more than simply presenting knowledge about threats and behaviours. Several psychological models of behaviour have established these antecedents of behaviour change. The authors evaluate the efficacy of different persuasion strategies, including the commonly used "fear appeals."

Case studies present current cybersecurity awareness campaign in the United Kingdom, Africa, Malaysia and Singapore. To investigate potential gaps represented in awareness campaigns, these four countries were chosen. Their vary not only in terms of cultural attributes, but also in the amount of money spent on promoting positive online behaviour.

4.1 CSA Campaign in UK

Different awareness campaigns are being carried out in the UK aimed at enhancing online security for companies and the public. The two most famous of these are presented below.

(i) The GetSafeOnline Campaign [32]: is a project sponsored jointly by several government agencies and the private sector, with a focus on home and business users. The upbeat message of "Get Safe Online" is intriguing in and of itself, emphasizing that people are ultimately responsible for staying safe online.
(ii) The Cyber Streetwise Campaign [33]: is focuses on consumers both at home and at work. These advise businesses to take five easy steps to strengthen their security.

The movement aims to change people's minds about online security by giving those tips and advice. "In short, the weakest links in the cyber security chain are you and me," says the campaign, which uses a constructive approach to influence user behaviour.

4.2 CSA Campaign in Africa

CSA campaigns in Africa establish two programme, as following;

(i) The ISC Africa [34]: is establish in An industry-wide and community-wide initiative to spread to Africa's people about the secure, in order to reduce the inherent risks and increase consumer trust. The campaign employs a positive messaging strategy to manipulate consumer behaviour in a more collectivist manner, with the slogan "Working together to ensure a secure online environment for everyone."
(ii) Parents' Corner Campaign [35]: The work carried out by government, industry and civil society is expected to be organized. Its goals are to guard children, parent empowerment, nurture children and build relationships and cooperation between the parties concerned.

Users protecting users in terms of their relationships and social group memberships are one of the key messages.

4.3 CSA Campaign in Malaysia

Government agencies, private sectors, and non-governmental organisations (NGOs) in Malaysia have launched national-level cyber safety awareness campaigns, including the Ministry of Women, Family, and Community Development (MWFCD), Malaysian Communications And Multimedia Commission (MCMC), CyberSecurity Malaysia (CSM), and Digi Telecommunications. The government-provided medium is critical for significantly raising awareness among especially to student and parent [28]. In 2016, MCMC launched a programme called Click Wisely, which hosted over 800 events for a growing 1.7 million people [36].

Malaysian Education Blueprint Higher Education 2015–2025 (MEBHE) suggesting Malaysian Education Blueprint Higher Education 2013–2025, which was released in 2013. Leveraging ICT to scale up quality learning across Malaysia is one of the changing pillars. Providing Internet access and a virtual learning experience to all 10,000 schools and optimizing the use of ICT for distance and self-paced learning expand access to high-quality teaching regardless of location or student skill level are the two of the proposed solutions [37].

4.4 CSA Campaign in Singapore

The initiative in Singapore has begun in schools, with the introduction of Cyber Wellness Education. Cyber Wellness Education is designed and implemented based on the needs of individual students and school environments. Cyber Wellness Education's primary aim is to provide students with lifelong social-emotional skills and values then can produce healthy, respect and responsible user, communication and technology [38].

5 Conclusion

Information protection may be jeopardised by cyber-attacks. As the rate of data use and Internet use continues to rise, cyber awareness has become increasingly important. Cybersecurity Awareness (CSA) knowledge is important to adapt in cyberspace usage especially among young generation. Society and authorities, by government support, have to play their roles in ensuring this generation from being trapped in the excessive disclosure of personal information. There is a need to devise a robust and realistic awareness program or events to ensure that the message is conveyed and that prevention is made possible by the young generation of Malaysia. Parents, colleges, workplaces, mass media, corporate bodies and associated agencies can all take part in implementing best practices when ensuring digital access to personal information. Early preparation for children to properly manage their online activities will instill patterns that will serve them well in adulthood, reducing their risk of victimization both early and later in life.

Acknowledgements This research is fully supported by UC TATI Short Term Grant (STG), 9001-2006, which help in completing the research in viable and efficient.

References

1. Muhirwe J, White N (2016) Cybersecurity awareness and practice of next generation corporate technology users. Issues Inf Syst 17
2. Hamad A, Irfan A, Jasim A, Yeusof A, Mohamad A, Ahmad A (2018) Understanding awareness of cyber security threat among IT employees. In: 2018 6th international conference on future internet of things and cloud workshops
3. Garba AA, Siraj M, Siti HO, Mahdi A (2020) A study on cybersecurity awareness among students in Yobe State University, Nigeria: a quantitative approach
4. Cox J (2012) Information systems user security: a structured model of the knowing—doing gap. Comput Hum Behav 28
5. Gabra AA, Sirat MB, Hajar S, Dauda BI (2020) Cyber security awareness among university students: a case study. Int J Adv Sci Technol 29
6. Michael PG, Albert NL (2008) Cyber security. Edward Elgar Publishing Limited, Cheltenham
7. Von SR, Van NJ (2013) From information security to cyber security. Comput Secur
8. Virginia DP, Anna DP, Sophie M, Emma P (2016) Cyberbullying among young people study. European Parliament.
9. Paulo DA (2018) 3 in 4 youngsters say they have been bullied online. https://www.channelnewsasia.com/news/cnainsider/3-in-4-teens-singapore-cyberbullying-bullied-onlinesurvey-10001480%0A
10. Nurul AA, Nooraini O (2019) Information privacy awareness among young generation in Malaysia. Inf Priv Awareness Young Gener Malaysia 5(2):1–10
11. Zwilling M, Ariel KG, Dušan Lesjak ŁW, Cetin F, Basim HN (2020) Cyber security awareness, knowledge and behavior: a comparative study. J Comput Inf Syst
12. Olasile BA, Emrah S (2020) Covid-19 pandemic and online learning: the challenges and opportunities. J Interact Learning Environ
13. Pranggono B, Arabo A (2020) COVID-19 pandemic cybersecurity issues. Internet Technol Lett
14. MCMC (2017) Melindungi Anak Daripada Ancaman Siber. Utusan Online (Rencana)
15. Hamid N (2017) Remaja, Kanak-Kanak Ketagihan Internet Serius. BH Online
16. Julian JJ, Surya N (2014) A survey of emerging threats in cybersecurity. J Comput Syst Sci 80(5)
17. Glancy DJ (1979) The invention of the right to privacy. Arizona Law Rev 21(1):1–39
18. Garner BA (2004) Black's law dictionary, 8th edn
19. MCMC (2020) MyCERT incident statistic
20. Brewster T (2016) Child tracker app "leaks 6.8 million texts, 1.8 million photos" from kids phones
21. Mathews L (2017) The latest privacy nightmare for parents: data leaks from smart toys
22. Singer N, Conger K (2019) Google is fined $170 million for violating children's privacy on YouTube
23. Rahim NHA, Hamid S, Kiah MLM (2019) Enhancement of cybersecurity awareness program on personal data. Malays J Comput Sci 32(3):221–245
24. Jun Z, Ge W, Carys D, Petr S, Julian ED (2019) 'I make up a silly name': understanding children's perception of privacy risks online
25. Maria B, Angela MS, Jason RCN (2019) Cyber security awareness campaigns: why do they fail to change behavior?
26. Rahman NAA, Sairi IH, Zizi NAM, Khalid F (2020) The importance of cybersecurity education in school. Int J Inf Educ Technol 10(5):378–382
27. Muniandy L, Muniandy B, Samsudin Z (2017) Cyber security behaviour among higher education students in Malaysia. J Inf Assur Cybersecur 1–13
28. Jambari N (2017) Parental awareness on cyber threats using social media. Malays J Commun 35(2)

29. Roepke R, Schroeder U (2019) The problem with teaching defence against the dark arts: a review of game-based learning applications and serious games for cyber security education. In: Proceedings of the 11th international conference on computer supported education
30. Quayyum F (2020) Cyber security education for children through gamification: research plan and perspectives. In: Extended abstracts—proceedings of the 2020 ACM interaction design and children conference, IDC 2020, Jan 9–13
31. Bada M, Sasse A, Nurse J (2014) Cyber security awareness campaigns: why they fail to change behavior. In: International conference on cyber security for sustainable society, p 11
32. GetSafeOnline Campaign (2021) www.getsafeonline.org
33. The Cyber Streetwise Campaign (2014) www.cyberstreetwise.com
34. ISC Africa (2013)
35. Parents Corner (2020)
36. MCMC (2017) Inisiatif-Inisiatif advokasi Suruhanjaya Komunikasi dan Multimedia Malaysia. http://www.klikdenganbijak.my/Landing-Page.aspx
37. Zulkifli Z, Molok NNA, Talib NHARS (2020) Cyber security awareness among secondary school students in Malaysia. J Inf Syst Digital Technol 2(2)
38. Ministry of Education S (2018) Cyber wellness: what is cyber wellness? https://www.moe.gov.sg/education/programmes/socia

Extending UTAUT to Understand the Acceptance of Queue Management Technology by Physicians in UAE

Adi A. AlQudah and Khaled Shaalan

Abstract Queue Management is significant to improve the provided healthcare services. Despite previous studies on technology adoption and users' intention to use various technologies in healthcare, users' acceptance of queue management solutions (QMS) have rarely been measured. End users in healthcare domain and their acceptance for information technologies are important to ensure the success of implementation for these technologies. It is essential to assess the level of acceptance for those users, and understand the related influencing factors. The objective of this study is to evaluate the factors that impact the acceptance of technology by physicians in UAE. It proposes a model based on the unified theory of acceptance and use of technology (UTAUT) extended by the construct "Trust". The study was conducted in healthcare organization in UAE and represents 63 physicians. To evaluate the proposed model, Structural Equation Modelling (SEM) was used, and data was analyzed using SmartPLS and SPSS Statistics Software. It was found that the proposed model could explain 62.3% of total variance in the behavioral intention to use Queue Management Solution (QMS) by physicians. The results also showed that Performance expectancy and Facilitating conditions are significantly influencing the physicians' intention to accept using QMS. Moreover, Trust as external factor has positive significant influence on the beliefs of physicians, especially when it comes to the importance of QMS to gain better performance results. This study can provide useful knowledge for decision-makers in healthcare organizations; who are planning to implement new QMS, or enhance the current available solution.

Keywords Technology acceptance · Unified theory of acceptance and use of technology · Structural equation modeling · Healthcare · Queue management

A. A. AlQudah (✉) · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, United Arab Emirates

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_77

1 Introduction

Over the years, managing the queues is common problem in healthcare and considered significant for the safety and overall satisfaction of patients [1]. Long queues in healthcare organizations can produce high levels of distraction for the employees instead of focusing on the original activities [2]. Queue management technology is becoming more popular to be implemented in healthcare organizations, to solve the issue of queues, gather data, and generate statistical reports about the current and future flow trends [3]. The acceptance to use queue management technology can be risky, and face resistance from health professionals, including physicians.

In the case of this study, the healthcare organization is looking to enhance the current implemented QMS, but there is a doubt regarding its benefit since there is a level of resistance from physicians to use the solution. Hence, the key question of this research is "what are the factors that can motivate the acceptance and actual use of QMS by physicians in UAE?". To answer such question, the study utilized the UTAUT model [4]. UTAUT model is the most relevant [5], and actively used in the technology acceptance studies in healthcare domain [5, 6]. The study considered the key constructs of UTAUT model, excluding the moderating factors. Also, the suggested model is extended by the construct "Trust" as inspired by the work of [5, 7]. Construct "Trust" is important to positively influence the beliefs of users, to obtain better performance results [7].

The key contribution of this study is to enrich the knowledge of research community, with respect to the application of UTAUT model to explain the acceptance of technology by physicians in UAE. The study can be also beneficial for decision-makers in healthcare, business analysts and IT corporations. It is important to find out how to enhance the level of technology acceptance (QMS in this case) in healthcare arena.

The coming sections will be responsible to give an idea about the related work, highlight the recent applications of UTAUT model in healthcare domain, present thorough explanation for the suggested model and its related hypotheses, report the conducted analysis and its based methodology. The final section will include the conclusion, main implications and limitations.

2 Literature Review

In 2003, the unified theory of acceptance and use of technology (UTAUT) has been proposed by [4], as a unified view for the technology acceptance by users. The authors could highlight various limits in other acceptance models, that were derived from Theory of Reasoned Action (TRA) [8]. Consequently, the TRA-driven models were combined to have a unified view to build UTAUT. It was confirmed that the ability to explain the intention to use technology by UTAUT is up to 70%, due to

the inclusion of facilitating conditions (FC) and social influence (SI) factors along with age, gender, experience and voluntariness as moderating factors [4].

The UTAUT model has been explored in different domain including banks [9], and education [10]. But it was claimed by [11] that there are limited applications of UTAUT to investigate information technology acceptance in healthcare.

In 2012, the acceptance of medical staff in Taiwan for an online patient-safety reporting system (PSRS) was studied by [12]. The study used the main constructs and moderating factors of UTAUT, with an integration for the factor of "Value of Perceived Consequence" as an attempt to make UTAUT more adequate. From 450 distributed surveys, only 183 (40.67%) responses were considered valid to be analyzed. The suggested model is significant and useful to predict the intention to use PSRS by medical staff. The results showed that the management support is significant to motivate the staff to use PSRS. The influence of performance expectance and perceived consequence on the intention to use PSRS was moderated by the age, experience, gender and occupation of staff.

Similarly, the UTAUT model was extended in [13] by perceived credibility and perceived organizational support, to understand the physicians' acceptance of E-prescribing in Pakistan. A total of 295 physicians participated in the study. The collected data were analyzed using structural equation modeling (SEM) assess associations among the constructs. All suggested hypotheses were supported, all key constructs of UTAUT along with perceived credibility had positive significant influence on behavioral intention to use the E-prescribing technology. As well, it was confirmed that higher perceived organizational support will lead to higher usage behavior of E-prescribing by physicians. In general, the proposed model could explain 56.1% of the variance in behavioral intention to use E-prescribing technology, and 40.6% in the variance in adoption of E-prescribing.

Using the original UTAUT model, [14] explored the factors that influence the acceptance of nutrition information system by nutrition officers in primary healthcare organization. The results of linear regression analysis for 50 valid responses, could show that the model explained 49.1% with contribution of performance expectancy, effort expectance, and social influence, and 50.9% is explained by other constructs.

3 Research Model and Hypotheses

The research model is adopted from the UTAUT model, including the key constructs with exclusion for the moderating constructs. UTAUT extended by adding the construct "Trust" because it is crucial to enhance the performance expectance (perceived usefulness of users) and their acceptance of technology within the healthcare domain [7].

Performance Expectancy (PE) is describing the point where an individual is confident, that using an information technology will be beneficial to improve the personal performance results [4, 15]. The work of [13, 15] could confirmed that

performance expectancy has positive influence on the user's behavioral intention in healthcare context. In this study, the following hypothesis has been developed:

H1: There is a significant positive impact for performance expectancy on behavioral intention of users while using technology in healthcare domain.

Effort Expectancy (EE) was described as the level of ease that is linked with suitability while using any system [4, 15]. Effort expectancy has the concept of using the system with no hassles [16]. It was verified that effort expectancy has positive impact on the users' behavioral intention in healthcare domain [13, 15]. Interestingly, the work of [17, 18] found that effort expectancy has positive influence on performance expectancy. So, the following hypotheses have been suggested in this study:

H2a: There is a significant positive impact for effort expectancy on the user's performance expectancy of technology in healthcare domain.
H2b: There is a significant positive impact for effort expectancy on behavioral intention of users while using technology in healthcare domain.

Social Influence (SI) refers to the perception of a person regarding the importance of social actors that expecting him/her to use technology [19]. It was confirmed by [20, 21] that perceived usefulness (performance expectancy) is impacted positively by social influence. Similarly, it was found by [13] that there is a positive effect for social influence on behavioral intention to use technology in healthcare. This study suggests the below hypotheses:

H3a: There is a significant positive impact for social influence on the user's performance expectancy of technology in healthcare domain.
H3b: There is a significant positive impact for social influence on the behavioral intentions (BI) of healthcare staff to use technology.

Facilitating Conditions (FC) is concerned about the existence of support and guidance from people including top management, guidance material and manuals to use information technologies [18]. The positive effect of facilitating conditions on healthcare staff's behavioral intention to use technology was verified in the studies of [13]. As well, [18, 22] could confirm that facilitating conditions is positively impacting the usage behavior of information systems in healthcare field Moreover, it was found that behavioral intention (BI) to use a technology has positive impact on the use behavior (UB) of that technology within the healthcare arena, regardless the type of technology [23, 24]. This study proposed the below hypotheses:

H4a: There is a significant positive impact for facilitating conditions on behavioral intention of healthcare staff to use technology.
H4b: There is a significant positive impact for facilitating conditions on healthcare staff's use behavior of technology.
H5: There is a significant positive impact for healthcare users' behavioral intention to use technology on the usage behavior of that technology.

Trust factor was studied as a predecessor for the acceptance of technology in different fields [19]. Trust has been defined as confidence of an organization and its services (level of reliability) [25]. Several studies have discussed that it is significant to trust the creator of a technology, and its positive impact on the intention to use that technology. Especially when the technology requires to allow share location feature, or digital exchanges [7]. As well, trust factor was underlined by [7] that it can positively influence the user's belief to gain better performance results, from using a particular technology within the healthcare context. This leads to the develop this hypothesis:

H6: There is a significant positive impact for user's trust of technology on the performance expectancy from that technology.

4 Research Methodology

Similar to previous studies that discussed the technology acceptance by healthcare professionals [13, 17, 26], this study have employed a quantitative approach to examine the developed hypotheses. The study included physicians who are working in healthcare organization in the United Arab of Emirates. A total number of 63 physicians to represent a target population of 88 physicians. The target population is known, so a non-probability sampling with convenience sampling techniques was used [27, 28]. Data were collected through survey that was conducted in February 2020, and sent to all prospective participants using their professional emails.

The survey contains measurements with 5-point Likert scale from "strongly disagree" (1) to "strongly agree" (5). Google forms tool [29] was utilized to build the survey and gather the research data. The total number of returned questionnaires was 63 (71.6%). All returned questionnaires were considered valid (No missing data) since all questions were mandatory to be answered. Krejcie and Morgan sampling method is common approach that is used to estimate sample sizes [30]. As per the work of Krejcie and Morgan in [31], the minimum accepted sample size is statistically calculated as S = 63 − 73 for population N = 75 − 90. In the case of this study, the sample size is 63, which is the minimum required sample size. Although it is inadequate sample, but it can be relatively acceptable due to the small target population, time of the survey (COVID19 outbreak), and the busy clinical schedule of participants.

The survey was divided to three sections, where first section included the ethics and consent section, and second section for the demographic details. The third section contained 28 different items to examine the developed hypotheses. Each item was extracted from previous related studies, and modified to the context of this study. Items of UTAUT constructs were mainly adapted from the work of [13, 20, 32–35], while the Trust items were extracted from [35, 36]. None of the items was related to specific personal details, or contact details to ensure the anonymous response from participants.

5 Findings

An analysis was conducted through the Structural Equation Modelling (SEM) using SmartPLS3 [37], along with and IBM SPSS Statistics Software v.25 [38]. The structural and measurement models (inner and outer) were examined. Final model and the hypotheses to be presented. The results, revealed that the model looks theoretically valid in order to explain the acceptance of technology (QMS) by physicians in UAE.

5.1 Demographic Details

Table 1 shows the demographic profile of the sample. Most of the respondents were males (71.43%), aged between 50 to 59 years old (42.86%), and mainly work within the Orthopedics clinic.

Table 1 Demographic details of participants

Sr.	Characteristics	Answer	Frequency	(%)
1	**Gender**	Female	18	28.57
		Male	45	71.43
2	**Age**	18 to 29	5	7.94
		30 to 39	15	23.8
		40 to 49	12	19.05
		50 to 59	27	42.86
		60+	4	6.35
3	**Clinic**	Orthopedics	28	44.44
		Internal medicine	15	23.8
		Family medicine	13	20.64
		ENT	7	11.11

5.2 Measurement Model

The measurement model represents the relations between indicators and latent constructs. The convergent and discriminate validities are essentials in order to assess the measurement model [39]. The convergent validity suggests the degree to which there is a high association between the constructs, that are theoretically identical. While, discriminant validity can provide to which degree that a specific construct is differentiated from other constructs [40].

As presented in Table 2, to examine the convergent validity, two methods were used. The loading for each measure, along with its associated constructs to assess

Table 2 Results of convergent validity

Constructs	Items	Loading	Cronbach's alpha	Composite reliability
Behavioral intention	BI1	0.869	**0.834**	**0.889**
	BI2	0.803		
	BI3	0.806		
	BI4	0.789		
Effort expectancy	EE1	0.775	**0.829**	**0.871**
	EE2	0.711		
	EE3	0.841		
	EE4	0.701		
	EE5	0.757		
Facilitating conditions	FC1	0.882	**0.904**	**0.933**
	FC2	0.888		
	FC3	0.880		
	FC4	0.877		
Performance expectancy	PE1	0.755	**0.817**	**0.872**
	PE2	0.764		
	PE3	0.702		
	PE4	0.832		
	PE5	0.827		
Social influence	SI1	0.819	**0.847**	**0.897**
	SI2	0.715		
	SI3	0.836		
	SI4	0.929		
Trust	TRU1	0.894	**0.885**	**0.925**
	TRU2	0.934		
	TRU3	0.862		
Use behavior	UB1	0.920	**0.873**	**0.922**
	UB2	0.866		
	UB3	0.893		

Bold results for Cronpach's Alpha means that they exceed the threshold value (0.70) and for theComposite Reliability means that the results exceed 0.80 and confirm good reliability

Table 3 Fornell Larcker scale for discriminant validity

	AVE	BI	EE	FC	PE	SI	TRU	UB
BI	0.669	**0.817**						
EE	0.576	0.611	**0.759**					
FC	0.776	0.665	0.617	**0.881**				
PE	0.579	0.688	0.497	0.534	**0.761**			
SI	0.687	0.349	0.598	0.408	0.219	**0.829**		
TRU	0.805	0.394	0.473	0.294	0.398	0.437	**0.897**	
UB	0.797	0.696	0.623	0.702	0.523	0.509	0.324	**0.893**

Bold diagonal items specify the square root of AVE scores that are higher than the associations amongthe constructs (the off-load diagonal items)

the internal consistency, while the second method to appoint the composite reliabilities [41]. Scales internal consistency was verified, since their Cronbach's Alpha exceeds threshold value (0.70) as suggested by [39]. Also, all constructs specified results of composite reliability that above 0.8 which confirmed good reliability [42]. Besides, the results of factor loadings for each item in the model's constructs were 0.70 or more. Thus, all items were valid, and can be added to the final study [43].

As seen in Table 3, the Average Variance Extracted (AVE) value is more than the threshold value of 0.50, which means that construct can create a minimum 50% of the variance of its items [43, 44]. On the other hand, Fornell-larcker scale and cross-loadings have been applied to examine if discriminant validity exists [44]. As in Table 3, bold diagonal items specify the square root of AVE scores, that are higher than the associations among the constructs (represented by the off-load diagonal items [45]).

5.3 Structural Model

The structural model is responsible to provide details about the predictive power of the model (R2) and path significance. PLS-Graph using a nonparametric test of significance (bootstrapping method) was conducted with 500 resamples to specify the significance levels, and path coefficients [46, 47]. The model's results as presented in Fig. 1 showed that it can explain 62.3% of total variance in the intention to use QMS by physicians in UAE, along with 58.6% for the actual use by physicians for the queue management technology. As well, all significant paths can be seen in Fig. 1.

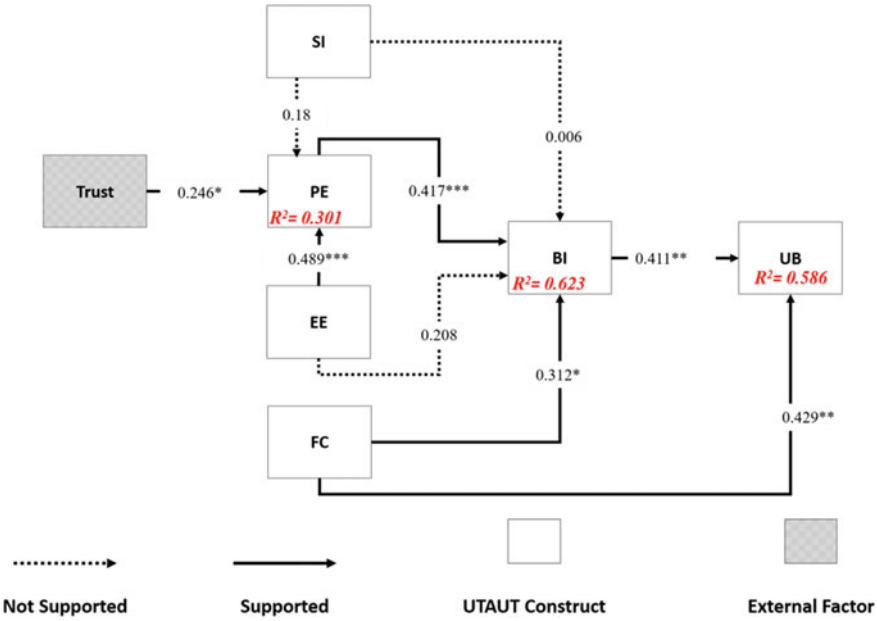


Fig. 1 Study final model—total effect hypotheses

6 Discussion

This study has explored the physicians' Intention to accept queue management technology in UAE. Association between BI and PE ($\beta = 0.417$, $P < 0.001$) as proposed in H1. This finding is consistent with the results of previous studies [15, 48], and indicates that a rise in performance expectancy can positively enhance physicians to accept and use QMS. Similarly, effort expectancy is significantly influencing performance expectancy ($\beta = 0.489$, $P < 0.001$), which is supporting hypothesis H2a, similar to what was reported by [17, 18]. This means that using QMS with hassle-free by physicians can improve their motivation to use the solution; since it is significantly helping them to do their work easily, and gain better results. But it was found that there is no significant effect from effort expectancy on the behavioral intention of physicians regarding the usage of QMS, where ($\beta = 0.208$, $P = 0.065$), so H2b was rejected. This finding is compatible with what has been found by [17, 49]. This finding can be explained by the experience level of the participants, who are familiar with using complex solutions (i.e., electronic medical records) to achieve their daily work activities.

Moreover, there was no significant path between social influence and performance expectancy ($\beta = 0.18$, $P = 0.085$), So H3a is not supported, which is opposite to the results of [21, 50]. Similarly, the results ($\beta = 0.006$, $P = 0.479$) of social influence shows non-significant impact on behavioral intention to use QMS

as per the settings. So H3b is not supported as well, which is similar to the findings of [51]. These results suggest that the physicians do not care about the opinion of their colleagues, to decide regarding the intention to use QMS, or enhance their belief that QMS can help to perform better. This might also refer to the relatively high levels of experience and age in the sample. Also, the results show significant influence for facilitating conditions on behavioral intention ($\beta = 0.312$, $P < 0.05$), then hypothesis H4a is supported, and compatible with the findings of [13, 52]. Likewise, H4b was confirmed ($\beta = 0.429$, $P < 0.001$), which is similarly was confirmed by [18, 22]. Having the right technical support and resources in place is significantly crucial, to motivate physicians' intention and actual usage of queue management technology. Furthermore, Fig. 1 shows that the actual use of QMS by physicians is significantly impacted by the behavioral intention to use QMS ($\beta = 0.411$, $P < 0.01$). this result supports H5 and similar to what has been found in [18]. Finally, the results indicated that the participated physicians trust QMS, to gain better performance results and achieve their work tasks proficiently, which is supporting H6 ($\beta = 0.246$, $P < 0.05$) and in line with the findings of [7, 19].

7 Study Implications

The study attempts to extend the UTAUT acceptance model, with the addition to "Trust" factor. Trust factor was compatible to be injected to the model, and it was helpful to strengthen its explanatory power with respect to the expected performance results. Practically, this study has presented a useful tool to managers, analysts, and decision-makers in healthcare organizations. They can identify the factors that are crucial to improve the acceptance to use QMS by physicians. The findings can help to ensure better future implementations of QMS, and can be employed to improve the current developed QMS by applying optimizations. Healthcare organizations and information technology providers need to consider the nature of users, gender and age.

8 Conclusion and Study Limitations

This study designed to recognize the factors that can motivate physicians in UAE to accept queue management technology. The data were collected through survey includes items with 5-point Likert scale from "strongly disagree" (1) to "strongly agree" (5). The analysis was completed using structural equation modelling (SEM). The results showed that that performance expectancy and facilitating conditions are significantly influencing the physicians' intention to accept and use QMS. Also, Trust was found significant for of physicians, to believe that QMS can help them to gain better performance results. In general, the proposed model could explain

62.3% of total variance in the intention to use QMS by physicians in UAE, along with 58.6% for the actual use of QMS.

The key limitations of the study are represented by small sample size (63 physicians), and the time of survey. The survey was distributed at the beginning stage of COVID19 outbreak in UAE (February 2020), and physicians were mainly busy with different activities that are related to the pandemic precautions. Also, the majority of respondents (more than 71%) were males which could bias results.

Acknowledgements This work is a part of a project undertaken at the British University in Dubai.

References

1. Davis MM, Heineke J (1994) Understanding the roles of the customer and the operation for better queue management. Int J Oper Prod Manag 14(5):21–34
2. Afolalu SA, Babaremu KO, Ongbali SO, Abioye AA, Abdulkareem A, Adejuyigbe SB (2019) Overview impact of application of queuing theory model on productivity performance in a banking sector. J Phys Conf Ser 1378(3):032033
3. Gosha KK (2007) Queueadmin: the advance queue management system for barbershop administration. Auburn University
4. Venkatesh V, Morris MG, Davis GB, Davis FD (2003) User acceptance of information technology: toward a unified view. MIS Q 27(3):425
5. Bennani AE, Oumlil R (2013) Factors fostering IT acceptance by nurses in Morocco: short paper. In: Proceedings—international conference on research challenges in information science
6. Perlich A, Meinel C, Zeis D (2018) Evaluation of the technology acceptance of a collaborative documentation system for addiction therapists and clients. Stud Health Technol Inform 247:695–699
7. Beldad AD, Hegner SM (2018) Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: a structural equation modeling approach. Int J Human-Computer Interact 34(9):882–893
8. Fishbein M, Ajzen I (1975) Belief, attitude, intention, and behavior: an introduction to theory and research. Addison-Wesley Pub. Co, Reading, Mass
9. AbuShanab E, Pearson JM, Setterstrom AJ (2010) Internet banking and customers' acceptance in Jordan: the unified model's perspective. Commun Assoc Inf Syst 26(1):493–524
10. Baltaci-Goktalay S, Ozdilek Z (2010) Pre-service teachers' perceptions about web 2.0 technologies. Procedia Soc Behav Sci 2(2):4737–4741
11. Holtz B, Krein S (2011) Understanding nurse perceptions of a newly implemented electronic medical record system. J Technol Hum Serv 29(4):247–262
12. Chang IC, Hsu HM (2012) Predicting medical staff intention to use an online reporting system with modified unified theory of acceptance and use of technology. Telemed e-Health 18 (1):67–73
13. Khan IU, Yu Y, Hameed Z, Khan SU, Waheed A (2018) Assessing the physicians' acceptance of E-prescribing in a developing country: an extension of the UTAUT model with moderating effect of perceived organizational support. J Glob Inf Manag 26(3):1–10
14. Masyarakat JK, Nurhayati S, Anandari D, Ekowati W (2019) Unified theory of acceptance and usage of technology (UTAUT) model to predict health information system adoption. KEMAS J Kesehat Masy 15(1):89–97

15. Ahmadi M, Mehrabi N, Sheikhtaheri A, Sadeghi M (2017) Acceptability of picture archiving and communication system (PACS) among hospital healthcare personnel based on a unified theory of acceptance and use of technology. Electron Physician 9(9):5325–5330
16. Davis FD, Bagozzi RP, Warshaw PR (1989) User acceptance of computer technology: a comparison of two theoretical models. Manage Sci 35(8):982–1003
17. Alaiad A, Zhou L, Koru G (2014) An exploratory study of home healthcare robots adoption applying the UTAUT model. Int J Healthc Inf Syst Inform 9(4):44–59
18. Ifinedo P (2012) Technology acceptance by health professionals in Canada: an analysis with a modified UTAUT model. In: 45th Hawaii international conference on system sciences, pp 2937–2946
19. Rajanen D, Weng M (2017) Digitization for fun or reward? A study of acceptance of wearable devices for personal healthcare. In: Proceedings of the 21st international academic Mindtrek conference, AcademicMindtrek 2017 January, vol 2017
20. Li J, Ma Q, Chan AH, Man SS (2019) Health monitoring through wearable technologies for older adults: smart wearables acceptance model. Appl Ergon 75:162–169
21. Briz-Ponce L, García-Peñalvo FJ (2015) An empirical assessment of a technology acceptance model for apps in medical education. J Med Syst 39(11)
22. Chang I-C, Hwang H-G, Hung W-F, Li Y-C (2007) Physicians' acceptance of pharmacokinetics-based clinical decision support systems. Expert Syst Appl 33(2):296–303
23. Dou K et al. (2017) Patients' acceptance of smartphone health technology for chronic disease management: a theoretical model and empirical test. JMIR mHealth uHealth 5(12):e177
24. Nematollahi M, Moosavi A, Lazem M, Aslani N, Kafashi M, Garavand A (2017) Factors affecting in adoption and use of electronic medical record based on unified theory of acceptance and use of technology in Iran. Shiraz E Med J 18(9):57582
25. Garbarino E, Johnson MS (1999) The different roles of satisfaction, trust, and commitment in customer relationships. J Mark 63(2):70
26. De Veer AJE, Peeters JM, Brabers AEM, Schellevis FG, Rademakers JJDJM, Francke AL (2015) Determinants of the intention to use e-health by community dwelling older people. BMC Health Serv Res 15(1)
27. Karno CG, Purwanto E (2017) The effect of cooperation and innovation on business performance. Qual Access Success 18(158):123–126
28. Ujang S, Rasyidha MH, Najib M (2020) The influence of intention to use medians potato varieties on the result of technological innovation development with technology acceptance models. J Manaj 24(1):38
29. Google (2021) Google forms: free online surveys for personal use. Google, 2020. Available: https://www.google.com/forms/about/. Accessed 03 Jan 2021
30. Chua LC, Penyelidikan J (2006) Sample size estimation using Krejcie and Morgan and Cohen statistical power analysis: a comparison. J Penyelid IPBL 7:78–86
31. Krejcie RV, Morgan DW (1970) Determining sample size for research activities. Educ Psychol Meas 30(3):607–610
32. Sarlan A, Ahmad R, Ahmad WFW, Dominic DD (2013) A study of SME private healthcare personnel acceptance of clinic information system in Malaysia. Int J Bus Inf Syst 14(2):238
33. Cimperman M, Makovec Brenčič M, Trkman P (2016) Analyzing older users' home telehealth services acceptance behavior-applying an extended UTAUT model. Int J Med Inform 90:22–31
34. ul Amin R, Inayat I, Shahzad B, Saleem K, Aijun L (2017) An empirical study on acceptance of secure healthcare service in Malaysia, Pakistan, and Saudi Arabia: a mobile cloud computing perspective. Ann Telecommun 72(5–6):253–264
35. Boonitt S (2019) Quality of health websites and their influence on perceived usefulness, trust and intention to use: an analysis from Thailand. J Innov Entrep 8(1)
36. Guo X, Zhang X, Sun Y (2016) The privacy-personalization paradox in mHealth services acceptance of different age groups. Electron Commer Res Appl 16:55–65
37. Ringle CM, Wende S, Becker J-M (2015) SmartPLS. https://www.smartpls.com/
38. SPSS® Statistics 25.0—overview | IBM. 2017. https://www.ibm.com/products/spss-statistics

39. Hair JF, Hult GTM, Ringle C (2016) A primer on partial least squares structural equation modeling (PLS-SEM)
40. Rahman IA, Memon AH, Abdullah NH, Azis AAA (2013) Application of PLS-SEM to assess the influence of construction resources on cost overrun. Appl Mech Mater 284–287:3649–3656
41. Ashill NJ, Jobber D (2010) Measuring state, effect, and response uncertainty: theoretical construct development and empirical validation. J Manage 36(5):1278–1308
42. Sekaran U, Bougie R (2016) Research methods for business: a skill building approach, 7th edn. Wiley
43. Hair JF, Risher JJ, Sarstedt M, Ringle CM (2019) When to use and how to report the results of PLS-SEM. Eur Bus Rev 31(1):2–24
44. Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. J Mark Res 18(1):39
45. Tsai JM, Cheng MJ, Tsai HH, Hung SW, Chen YL (2019) Acceptance and resistance of telehealth: the perspective of dual-factor concepts in technology adoption. Int J Inf Manage 49:34–44
46. Gil-Garcia JR (2011) Using partial least squares in digital government research. In: Handbook of research on public information technology. IGI Global
47. Vinzi VE, Trinchera L, Amato S (2010) PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement. In: Handbook of partial least squares. Springer, Berlin, Heidelberg, pp 47–82
48. Idoga PE, Toycan M, Nadiri H, Çelebi E (2019) Assessing factors militating against the acceptance and successful implementation of a cloud based health center from the healthcare professionals' perspective: a survey of hospitals in Benue state, northcentral Nigeria. BMC Med Inform Decis Mak 19(1)
49. Lárusdóttir M, Cajander Å, Gulliksen J (2014) Informal feedback rather than performance measurements—user-centred evaluation in Scrum projects. Behav Inf Technol 33(11):1118–1135
50. Becker D (2016) Acceptance of mobile mental health treatment applications. Procedia Comput Sci 58:220–227
51. Francis RP (2019) Examining healthcare providers' acceptance of data from patient self-monitoring devices using structural equation modeling with the UTAUT2 model. Int J Healthc Inf Syst Informatics 14(1):44–60
52. Kim S, Lee K-H, Hwang H, Yoo S (2016) Analysis of the factors influencing healthcare professionals' adoption of mobile electronic medical record (EMR) using the unified theory of acceptance and use of technology (UTAUT) in a tertiary hospital. BMC Med Inform Decis Mak 16(1):12

The Mediating Effect of Competitive Advantage on the Relationship Between Human IT Resources and Performance

Ehsan Ali Salem Alqararah and Wan Mohd Nazri Bin Wan Daud

Abstract Human IT capability gained mainly from the organization based on the strategic perspective of human resources. Human capital is a valuable resource for organizations as intellectual assets. Hence, human IT resources can be perceived as a strategic capability that directly impacts performance in the Jordanian banking sector. This research aims to examine the mediating effect of competitive advantage between IT human resources on performance in the Jordanian banking sector from 2014 to 2019. To achieve the research objectives, it is expected from this study to generate practical knowledge that can be used by the IT human resources of the respective banks to create a competitive advantage and improve their performance. Additionally, this study is expected to add to knowledge in various ways, including demonstration of how IT human resources as a dynamic resource could impact the competitive advantage and firm performance in light of the Resource-Based View (RBV) theory. Thus, a quantitative method is used to collect and analyze data from 16 Jordanian banks; the targeted population is about 853 management staff of the Jordanian banks. The empirical results showed the mediating effect of competitive advantage on the relationship between IT human resources and performance; this conclusion related to the sufficient capturing of the respective setting that in turn becomes a positive impact. Furthermore, human IT resources in the given context could be revealed with other resources and capabilities, or through different business routines.

Keywords IT human resources · Performance · Competitive advantage · Resource-based view (RBV)

E. A. S. Alqararah (✉) · W. M. N. B. W. Daud
Faculty of Business and Management Science, Universiti Sultan Zainal Abidin, Kuala Nerus, 21300 Terengganu, Malaysia
e-mail: Wanmnazri@unisza.edu.my

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_78

1 Introduction

In the present century, expenditures on Information Technology (IT) accounts for a large portion of organizations' discretionary spending as businesses are continually searching for ways to leverage IT to achieve a competitive advantage [1–3]. Human capital can be easily accessed, and firms are striving to exploit the available technologies to improve their performance. The Resource-Based View (RBV) theory has been used by various scholars to examine the IT elements that have been suggested to guarantee a sustained competitive advantage [4–6]. Despite the considerable research attention provided since the early 1980s to the strategic use of IT, information system (IS) researchers are yet to arrive at a consensus on whether IT elements can provide a sustained competitive advantage [7].

IT human resources as one of the IT capabilities components seem to be the new key to competitive advantage [8]. Many scholars have reasoned that the development of IT human resources in accordance with the business design and strategy is becoming the cornerstone of competitive advantage [9]. The RBV theory of competitive strategy has been considered by many scholars as a platitude [10, 11]. However, the RBV theory has been mostly criticized based on the superficial and indirect way of its evaluation. However, only few studies have explained issues regarding the accurate detection of the existence of significant resources and capabilities particularly in correlation to the impact of superior performance. Likewise, there are limited studies that explain the mediating effect of competitive advantage between any components of IT capabilities and performance [3]. Similarly, findings regarding the effect of human IT resources on the performance are almost inconclusive. The contradictory arguments of IT scholars imply that IT components do not necessarily lead to superior performance [2, 3]. This study strived to fill this gap that arises from the lack of studies examine the mediating effect of competitive advantage between specific IT component and performance, particularly in the developing country [12]. By covering the background of this study, the remaining paper structure is organized into four sections. Literature review is reviewed in section two; section three described methodology used in this study, followed by the results and discussions under section four. At the end of this paper, section five presents conclusions and future directions.

2 Literature Review

In the RBV theory, Barney [4] interpreted the organizational human resources component as experience and training, as well as the attitudes and relationships of the employees of a corporation. Wright et al. [13] examined human resources and found it primarily necessary to the creation of a high-resource pool of human resources from the current or new resources. The scholars pointed out that this pooling would then gain an advantage that would increase the company's

performance. By being valuable, rare, and inimitable, it is organized by an organization to capture the value of resources in line with the business strategy.

The literature on human IT resources has mainly concentrated on IT management skills. The initial attempt to a systematic highlighting of the requisite skills and expertise of IT professionals was made by Lee et al. [14] and Rockart et al. [15]; they attempted to clarify a wide range of experience and knowledge, such as understanding business issues, perceiving challenges, technical skills, as well as interpersonal skills.

Bharadwaj [8] further evaluated the aforementioned relationship and its evidence and extended the framework by adding (i) the ability to more efficiently incorporate business and IT planning processes than other businesses, (ii) the ability to create cost-effective and efficient business applications faster than rivals, (iii) the ability to connect and work better with organization units, and (iv) the opportunity to predict the company's future market needs and to innovate precious new product features ahead of rivals. It should be recognized that such contemporary research studies agree that organizations need a wide range of skills to provide for their company and to improve performance[16–18]. This prompted the claim by Bharadwaj [8] that organizations with strong human IT capabilities can achieve higher levels of performance. Minbeava et al. [19] argued that practices enhance the value of IT business and sustainable competitive advantage can be gained through exceptional IT management skills. The following hypotheses have been formulated basis on the proposed statements from the research works described above:

H1: Human IT resource has a positive impact on competitive advantage in the Jordanian banking sector.
H2: Human IT resource has a positive impact on performance in the Jordanian banking sector.

Several studies have concluded that business processes can be more successfully implemented by organizations with good IT practices [19]. These practices applied by human IT capital to build reliable, productive applications that serve the company's business needs. Mikalef et al. [20] assessed a firm's future business needs and established the need for valuable new product characteristics before competitors to generate a sustainable competitive advantage.

DiClaudio [9] suggested that businesses should understand the need to concentrate more on how to handle human skills in their IT alignment, as many organizations have become reliant on IT, and adds that it is essential to increase the value of the behaviors of IT staff in the organization and recognize the ability of people generate new creative ideas. Joseph et al. [17] pointed to the need for a broad view of human skills that guides the relationship between competitive advantage and human IT resources. Bharadwaj [8] confirmed this hypothesis to indicate that superior human IT capital may lead to competitive advantage.

Based on the RBV, IT researchers have proposed that organizations should strive to differ from their competitors based on the IT resources of their business.

A powerful human IT capital is a strong organizational capability; likewise, organizations aim to employ a strong human capital stock for their human IT, which is necessary for improved performance, but may not be sufficient [21]. In predicting performance, this study proposed the following hypotheses regarding the relationship between these variables:

H3: Competitive advantage mediates the relationship between human IT resources and performance in the Jordanian banking sector.

3 Methodology

To address the research problem, survey-based research was conducted using a 5-point Likert scale. This research is explanatory as it relies on the use of the existing theories to find potential solutions to a specific business problem in a new context [22]. The adopted research approach is quantitative method in this study [23]. Since this relationship between the variables had to be examined, a correlational research method was utilized. In line with the existing studies in this field, a cross-sectional analysis was used in this study, meaning that data collection was done at one point in time only.

The study targeted a population of approximately 853 managers across all Jordanian banks. Two publicly accessible databases (the ABJ Association of Banks in Jordan and the Jordan Central Bank) were used in this study as both associations consented to the use of their databases in this study. A representative sample of 270 respondents was selected for the data collection based on their perceptions of the research subject. The targeted individuals in the abovementioned banks were representative business managers of these banks; they were required to rank their firms in terms of human IT resources and performance. This approach aims to ensure that the data is generated as randomly as possible. The measurement instruments used in the previous empirical studies were used for all the constructs used in this study.

Data analysis and hypotheses tests were done using Structured Equation Modeling (SEM) and Partial Least Squares (PLS) techniques as PLS remains the most appropriate tool used for statistical analysis in this field. Several data analysis steps were performed to ensure the accuracy of the statistical findings.

4 Results and Discussions

This research employed Structural Equation Modeling (SEM) approach to examine the impact of human IT resources on competitive advantage and performance in the Jordanian banking sector from 2014 to 2019. The partial Squares method (PLS) was

applied to the cross-sectional data that related to the indicators of human IT resources, competitive advantage, and performance during the study period. The study was performed based on the perceptions of the managers of the Jordanian commercial banks and other relevant studies previously conducted on banking performance and competitiveness in Jordanian banks and other countries around the world.

The validation of the measurement model involves several statistical tests that include tests for convergent validity, internal consistency reliability, and discriminant validity [27, 28]. Tests for internal consistency reliability were based on the value of the Composite Reliability (CR) and Cronbach's alpha (α). Table 1 showed the calculated α values of all the constructs in this study; observably, all the values were between 0.938 and 0.943, whereas the CR values were between 0.945 and 0.951. These values were more than the recommended 0.7 benchmark for α and CR values [27, 28] (see Fig. 1).

The convergent validity values of all the constructs were calculated and presented in Table 2. The values were based on checking the Average Variance Extracted (AVE) and factor loadings of each item for all the constructs. For all the items, it is expected that the factor loadings should be >0.70 [27, 28]. Table 2 showed that all the items achieved factor loading values of >0.708 except CA4 and CA6 which achieved factor loading values of 0.695 and 0.692, respectively; hence, both items were removed from the measurement model. The measurement model was also tested for CV; as per [27, 28], each construct in the measurement model

Table 1 The measurement instruments of the constructs

Construct	Source	Items in survey
Human IT resources	Adopted from: Bharadwaj [8]	Excellence in work and communication within business divisions Ability to promote reliable applications that support the business needs more effectively than competition Appropriate degree of integration of IT and business planning processes effectively Excellence in managerial skills and experience in the IT divisions Excellence in technical skills and experience in the IT divisions Excellence in IT project management practice, ability to predict future business needs of the organization and adopt valuable new product \service features before competitors
Performance	Adopted from: Capon et al. [24]	Market-based performance
	Adopted from: Powell and Dent-micallef [25]	Profitability
Competitive advantage	Porter [26]	Cost, quality, time, innovation

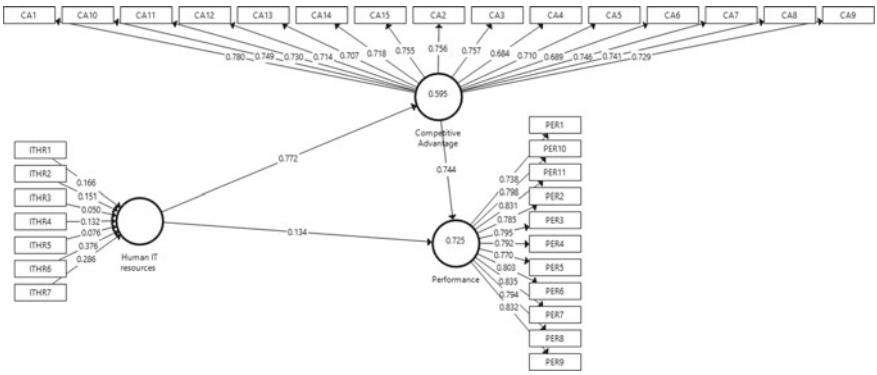


Fig. 1 The measurement model assessment

Table 2 Validity and reliability indicators

Construct	Item	Factor loading	CA	CR	AVE
Competitive advantage	CA1	0.780	0.938	0.945	0.535
	CA10	0.749			
	CA11	0.730			
	CA12	0.714			
	CA13	0.707			
	CA14	0.718			
	CA15	0.755			
	CA2	0.756			
	CA3	0.757			
	CA4	0.684			
	CA5	0.710			
	CA6	0.689			
	CA7	0.746			
	CA8	0.741			
	CA9	0.729			
Performance	PER1	0.738	0.943	0.951	0.637
	PER10	0.798			
	PER11	0.831			
	PER2	0.785			
	PER3	0.795			
	PER4	0.792			
	PER5	0.770			
	PER6	0.803			
	PER7	0.835			
	PER8	0.794			
	PER9	0.832			

Table 3 Formative construct assessment

Construct	Item	Outer weights	t-values	VIF
Human IT resources	ITHR1	0.166	1.962	2.561
	ITHR2	0.151	1.936	2.434
	ITHR3	0.050	0.733	2.048
	ITHR4	0.132	1.588	2.713
	ITHR5	0.076	0.939	2.541
	ITHR6	0.376	4.504	2.706
	ITHR7	0.286	3.573	2.165

must achieve an AVE value closer to or >0.5. From Table 2, it was also evident that all the constructs in the measurement model achieved good CV and reliability.

To ensure no multicollinearity in the model, Variance Inflation Factor (VIF) values below 3.3 are considered acceptable in determining low multicollinearity. Based on the results in Table 3, the VIF values were found to be lower than the recommended threshold of 5 [28]. The t-values must also be >1.64 at a 5% significance level to establish significant outer weights. The outer weights in formative measurement models are significantly different from zero; the t-value assesses the significance of each indicator weight; hence, all the formative indicators were retained for further analysis.

The next step is the assessment of the structural model; this was done using the Bootstrapping technique with a resample of 5,000 employed to validate the developed measurement model. For the research model to be considered valid, the model must achieve a path coefficient and p-value of <0.05, while the t-value must be >1.96. From Table 4, all the path coefficients of the model were significant (t-value > 1.96; p-value < 0.000).

The mediation effect exists when the path coefficient between Human IT resources (exogenous construct) and the competitive advantage (mediating construct) is $\neq 0$, and if the path coefficient between the competitive advantage (mediating construct) and performance (endogenous construct) is $\neq 0$. In addition, the path coefficient between Human IT resources and performance is ≥ 0 and smaller than the path coefficient (between exo. construct and endo. construct) without exist the mediator in the same model (see Fig. 2).

Table 4 Path coefficient

Hypotheses	Relationship	Coefficient	t value	P values	Results
H1	Human IT resources → competitive advantage	0.772	23.493	0.000	Supported
H2	Human IT resources → performance	0.134	2.028	0.021	Supported
H3	Human IT resources → competitive advantage → performance	0.744	9.745	0.000	Supported

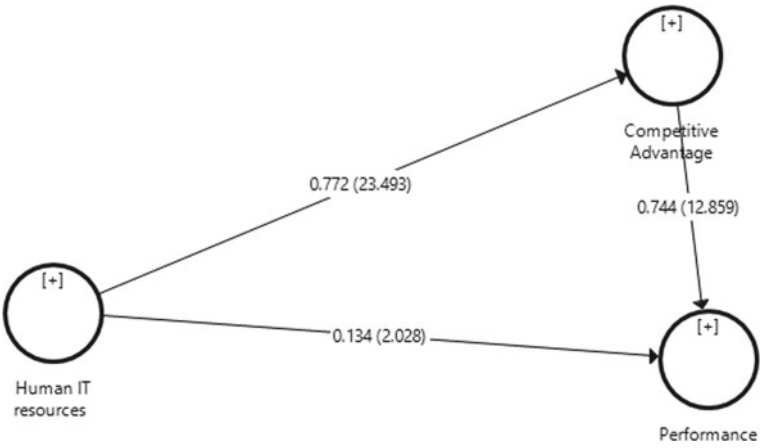


Fig. 2 Path coefficient values for constructs

5 Conclusion and Future Directions

The measured impact remained positive and the construct could not be disregarded. An explanation of this phenomenon may be the adequate capture of the respective environment which in turn has a positive effect. Another interpretation could be revealed along with other resources and skills, or through various business routines of human IT resources in the context given.

This study showed that many qualifications have become significant for the respective industry, such as the ability to anticipate future business needs, implement reliable applications that support such needs, the ability to function and interact independently within business segments, and the ability to more efficiently integrate IT and business planning processes.

The mediating role of competitive advantage on the relationship between IT human resources and performance was established in this study. This relationship implies that organizations deploy human IT capabilities with superior skills and well-qualified focus on its major areas; this plays a significant role in improving the level of performance and indicates that elements other than technical IT skills and managerial IT skills are important. Such skills are those that focus on business unit incorporation, such as anticipation of future business needs and integration of IT with business units and processes management. The element of organizational routines and the model by [29] can be used as a basis in such research work; for instance, more research work could focus on the mediating variables between human practices and organizational performance. Finally, scholars and practitioners could benefit from studies on the generalization of the results of this study since it focused only on one developing country.

References

1. Madon S, Krishna S (2018) The digital challenge: information technology in the development context. Routledge
2. Sampler JL, Earl MJ (2014) Whats your information footprint? MIT Sloan Manag Rev 55 (2):95
3. Mithas S, Rust R (2016) How information technology strategy and investments influence firm performance: conjecture and empirical evidence. MIS Q 40(1):223–245
4. Barney J (1991) Firm resources and sustained competitive advantage. J Manage 17(1):99–120. https://doi.org/10.1177/014920639101700108
5. Grant RM (1991) The resource-based theory of competitive advantage: implications for strategy formulation. Calif Manage Rev 114–135
6. Wernerfelt B (1984) Harmonised implementation of application-specific messages (ASMs). Strateg Manag J CINCO 2:1–12. http://onlinelibrary.wiley.com/doi/10.1002/smj.4250050207/abstract
7. Haseeb M, Hussain HI, Kot S (2019) Role of social and technological challenges in achieving a sustainable competitive advantage and sustainable business performance. Sustainability 11(14):3811
8. Bharadwaj AS (2000) Research article a resource based perspective on information technology capability and firm performance. MIS Q 24(1):169–196
9. DiClaudio M (2019) People analytics and the rise of HR: how data, analytics and emerging technology can transform human resources (HR) into a profit center, vol 18, no 2, p 5
10. Powell TC (2001) Competitive advantage: logical and philosophical considerations. Strateg Manag J 22(9):875–888. https://doi.org/10.1002/smj.173
11. Ashour ML (2019) Schools thought of competitive advantage. Int J Bus Manag Invent 7(2):01–07
12. Arora B, Rahman Z (2015) Article information: information technology capability as competitive advantage in emerging markets: evidence from India. https://doi.org/10.1108/IJoEM-07-2015-0127
13. Wright PM, Mcmahan GC, Mcwilliams A (1994) Human resources and sustained competitive advantage: a resource—based perspective center for effective organizations. Int J Hum Resour Manag 0–34
14. Lee DMS, Trauth EM, Farwell D (1995) Critical skills and knowledge requirements of IS professionals: a joint academic/industry investigation. MIS Q Manag Inf Syst 19(3):313–337. https://doi.org/10.2307/249598
15. Rockart JF, Earl MJ, Ross JW (1996) The new IT organization: eight imperatives. Cent Inf Syst Res 38:43–55
16. Crook TR, Todd SY, Combs JG, Woehr DJ, Ketchen DJ (2011) Does human capital matter? A meta-analysis of the relationship between human capital and firm performance. J Appl Psychol 96(3):443–456. https://doi.org/10.1037/a0022147
17. Joseph D, Ang S, Chang RHL, Slaughter SA (2010) Practical intelligence in IT: assessing soft skills of IT professionals. Commun ACM 53(2):149–154. https://doi.org/10.1145/1646353.1646391
18. Alpkan L, Bulut C, Gunday G, Ulusoy G, Kilic K, Gunday G (2010) Organizational support for intrapreneurship and its interaction with human capital to enhance innovative performance. Manag Decis 48(5):732–755
19. Minbaeva DB, Shell RD (2017) Building credible human capital analytics for organizational competitive advantage. Hum Resour Manage 57(3):701–713. https://doi.org/10.1002/hrm.21848
20. Mikalef P, Pateli A, Van De Wetering R (2020) IT architecture flexibility and IT governance decentralisation as drivers of IT-enabled dynamic capabilities and competitive performance: the moderating effect of the external environment article history. Eur J Inf Syst 1–29. https://doi.org/10.1080/0960085X.2020.1808541

21. Barnes E, Adam C (2018) The importance of human resources in a globalised economy: a conceptual framework. Can J Appl Sci Technol 5(2)
22. Neuman W (2003) Social research methods, qualitative and quantitative approaches, 5th edn. Allyn & Bacon, Boston/US
23. Leedy D, Ormrod J (2005) Practical research—planning and design, 8th international. Pearson Education, New Jersey/US
24. Capon N, Farley JU, Lehmann DR, Hulbert JM (1992) Profiles of product innovators among large U.S. manufacturers. Manage Sci 38(2):157–169. https://doi.org/10.1287/mnsc.38.2.157
25. Powell TC, Dent-Micallef A (1997) Information technology as competitive advantage: the role of human, business, and technology resources. Strateg Manag J 18(5):375–405. https://doi.org/10.1002/(sici)1097-0266(199705)18:5%3c375::aid-smj876%3e3.3.co;2-z
26. Porter ME (2008) Competitive advantage: creating and sustaining superior performance. Simon and Schuster
27. Ramayah T, Cheah J, Chuah F, Ting H, Memon MA (2018) Partial least squares structural equation modeling (PLS-SEM) using SmartPLS 3.0: an updated guide and practical guide to statistical analysis, 2nd edn. Pearson, Kuala Lumpur, Malaysia
28. Hair Jr JF, Hult GTM, Ringle C, Sarstedt M (2016) A primer on partial least squares structural equation modeling (PLS-SEM). Sage publications
29. Grant RM (1996) Towards a knowledge-based theory of the firm. Strateg Manag J 17 (Winter):109–122

A Comprehensive Analytical Case Study on Plisse Fashion Design L.L.C

Mohamed Abdulla, Omar Khalid, Said A. Salloum, and Khaled Shaalan

Abstract Knowledge management is a fundamental process and an imperative framework on which organizations achieve competitive advantages. Companies realize that to benefit from the wide range of knowledge available, they must find ways of managing organizational knowledge. However, how companies manage their knowledge is dependent on the activities performed within the organization. Proper planning and implementation of a knowledge management tool allow the company to achieve its objectives while increasing the profit margin. This report provides an analysis of knowledge management in the Plisse Fashion Design Company. The study utilizes interviews as its primary data collection approach. In addition, by carefully examining the accomplished and most distinctive work as well as closely analyzing Plisse's particular case and current knowledge management practices, the study produces various potential areas of improvements. The recommended areas come from the literature review and analyses of case studies. The significant areas of knowledge management are planning and workflow management. Organizations are advised to establish plans and useful tools to exploit the existing knowledge in their departments fully. Furthermore, Plisse should involve its employees and consider the cultural implications of the implementation of knowledge management to ensure it leads to the anticipated results.

Keywords Knowledge management (KM) · Knowledge sharing (KS) · Knowledge transfer (KT) · Knowledge management tools · Knowledge management systems

M. Abdulla · O. Khalid · S. A. Salloum (✉) · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, United Arab Emirates
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Research Institute of Sciences and Engineering, University of Sharjah, Sharjah, United Arab Emirates

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_79

1 Introduction

1.1 Overview and Objectives

This paper aims at analyzing a technology, which facilitates the sharing and management of knowledge in a business organization. The study's recommendations will be based on a pre-executed literature examination that covers a total of 11 organizational knowledge management investigations. In addition, the paper evaluates how such organizations can benefit from the principles of knowledge management. Besides, the paper demonstrates the usefulness of knowledge management in human resources and precisely how it facilitates workflows in a human resources department.

1.2 Method and Approach

The report consists of four main sections; first, we provide an overview of the background of the organization and their practice of knowledge management. Then, the paper will give an overview of research papers deemed useful in this study. After that, we will identify the main issues of Plisse, and it will provide solutions to enhance the sharing of knowledge in the organization.

1.3 The Research Questions

The study will attempt to approach and discuss the primary research question concerning the major areas of improvements that can be embraced by Plisse and similar organizations to enhance their knowledge management practices and outcomes. Additionally, the paper will discuss various secondary research question regarding the effectiveness of Plisse's current knowledge management approach and tools, and how can they be furtherly improved to better serve and assist in accomplishing the company's future vision and ambitions.

1.4 The Interview (Data Collection)

The paper relied on information from several semi-structured interviews with some of Plisse's employees. We contacted the interviewees through telephones and through face-to-face interviews concerning the issue of knowledge management, the current application of the system, constraints, opportunities, and specific discussion about the knowledge management tools they are currently using. Prior to

the interview, we informed the participants about the topic and objectives of the research through phone calls. Interviewees were nominated and selected based on their role and involvement in decision making and knowledge management within the organization. Thereafter, the collected data was analyzed qualitatively using a thematic analysis approach with respect to the findings in the literature.

2 Organization Background (The Case)

This study will focus on the human resources department in Plisse Fashion Design Limited Liability Company, a highly ambitious and fast-growing fashion design company that studies the latest fashion trends and transform them into a final clothing product as well as footwear, textiles, and other fashion-related accessories. The company is based in Dubai, United Arab Emirates. Plisse company was founded in 2007 as a textile trading company that imports high-quality textiles and resell them to other fashion design companies. The company started with only 20 employees. Now, the company has around 200 employees across the city of Dubai and planning to expand to other cities in the country and possibly to other countries within the Gulf area in the upcoming decade. Human resources is one of the fundamental departments in a company that is growing fast and planning to massively expand by hiring a large number of employees within a short period of time. Thus, innovation and proper planning are essential to achieve optimum workflow. The department is responsible for planning, managing, and initiating innovation at the company. Due to the increasing competition in the fashion industry, Plisse continuously introduces new trends and ideas, and the department is undergoing rigorous changes to equip its employees with relevant knowledge. Therefore, it is critical for Plisse company to implement practical tools that will utilize knowledge in a way that provides insight and offer solutions to existing problems. Besides, Plisse needs to integrate the past and current experience in a manner that is understandable to the employees. Capturing the past and existing knowledge provide a platform over which Plisse can plan better about its projects and learn from past mistakes. By providing new insights to the employees, they will improve performance and become more efficient.

2.1 Knowledge Management in the Selected Organization

In the beginning, Plisse did not even have a human resources department as they did not see the need for it when the startup was first founded, not to mention that they have never considered using a knowledge management tool. The founders did not foresee the hidden possible potentials of such a tool, considering the minimal number of employees they had back then. As the company kept on rapidly growing with multinational employees coming from various backgrounds and knowledge,

Plisse executives understood the necessity of such a tool to manage, organize, and share this massive amount of knowledge within the company. Currently, Plisse uses a custom-built knowledge management tool that was designed and built by its IT department based on the company's specific needs. The tool stores and manages the information that facilitates the utilization of knowledge to meet the organization's goals. Some of the components used in Plisse include wikis, work pad documents, internal chat links, bulletin boards, among others. These are some of the web-based tools, and they are controlled and managed from a central platform, which is their own KM tool. These tools contain data, which relates to each unit, such as project teams and other essential practices in Plisse. Plisse's KM tool ensures that knowledge is available in the organization for both employees and managers to make projections of future activities and projects. The KM tool also utilizes the on-the-job training (OJT), where an experienced person with a specific set of skills trains other employees. Plisse uses its KM tool as a valuable source of information and helps managers tackle the many issues that experience when completing different projects. The wikis used are useful because employees access information in a more centralized manner hence adding more new information to the departments accordingly. During the interviews, the line and design managers stated that wikis are beneficial to the company because they also provide information on the environment, and the tool is more specific as it gives troubleshooting solutions for operational errors.

2.2 Knowledge Management in the HR Department

The human resources manager noted that technology offers opportunities in the department by providing more effective ways of handling employees, assignment of employees, production of payrolls, and better ways of addressing their issues. Generally, technology has become a focal point in the human resources department for improving the welfare of employees. As technology advances, the market of fashion design as customers utilize the digital devices tool to browse and purchase products. An increase in demand calls for better projects' management and more efficient ways of handling employees. Besides, more efficient methods of sharing and distribution of information were required, and the human resources department initiated the web-based initiatives that created structured approaches for standardizing and controlling information accessed from the internet concerning the operations and performance of the employees.

The web-based initiatives played a significant role in information sharing and provided a functional infrastructure for knowledge distribution [1–4]. An essential component in the web-based technique is its web portal, which is a type of expertise web-based competence boosting the relationship between Plisse and the employees. In particular, the web portal provides information about the company and what they are expected to do and promotes interaction with the team managers. The web portal creates a "playground" for enhancing their networking skills, and gain more

insights about the company. Plisse developed its web portal with the idea of knowledge networking between Plisse's different offices in Dubai. As a future plan, the technology is to be used among their upcoming branches and offices in Abu Dhabi and Sharjah, as well as their planned offices in Saudi Arabia, Kuwait, and Bahrain.

2.3 Selected Research Papers and Topic

This paper identified 11 papers, which focused on the concept of management. These papers explained what knowledge management involves and the benefits of the technique to an organization. One of the papers reviews the application of knowledge management in companies such as Apple Inc., Siemens, NASA software engineering, among others [5]. The second paper, A Framework for Knowledge Management and Automated Reasoning, demonstrates the application of knowledge management in human resources management and how it encourages innovation at Ericsson [6]. The third article, Semantic Knowledge Management System to Support Software Engineers, is knowledge management and sharing within project teams, with a focus on a qualitative study in IT organizations [7]. Other papers included in this research are:

- Knowledge management in human resources management: Foreign-owned subsidiaries' practices in four CEE countries by [8].
- Sustainable Knowledge Management and Its Impact on the Performances of Biotechnology Organizations by [9].
- Knowledge Management and Job Satisfaction. International Journal of Academic Research in Business and Social Sciences by [10].
- Knowledge management practice in the private sector: Building the way for Saudi Arabia's strategic growth and transformation to a knowledge-based economy by [11].
- The role of the knowledge management process and intellectual capital as intermediary variables between knowledge management infrastructure and organization performance by [12].
- Impacts of knowledge management on innovations in higher education institutions: An empirical evidence from Vietnam by [13].
- The impact of knowledge management on organizational productivity: A case study on Koosar Bank of Iran by [14].
- Knowledge management process model. Entrepreneurship and Sustainability Issues by [15].

These papers will demonstrate the application of knowledge management in other multi-sector companies, and factors to consider when improving the management systems. The articles provide rich literature on how Plisse can adjust their policies to ensure that the management of employees is more efficient.

2.4 Summary

In this segment, we have provided an overview of the Plisse company concerning knowledge management. Besides, this section has identified the major papers, found useful for this project. The following part will evaluate methods of knowledge management at Plisse and discuss any problems and possible solutions.

3 Major Problems Identifications and Analysis

This part will focus on the issues arising from the application of knowledge management tools at Plisse. Besides, it will discuss the alternative and solutions based on the selected papers.

3.1 Major Problems and Analysis

According to the human resources manager, the general reflection about the web-based methods of knowledge management was productive meeting the needs of the organization. From the information obtained from the human resources manager, the local wiki system used by the employees is simple and convenient for the people who had an experience. The on-the-job training method was also useful, especially for the new employee. Again, putting everything on the web ensured that every employee accessed the information, which enhanced coordination within the organization. However, there are several issues related to the use and implementation of web-based techniques.

1. The on-the-job training method was more appropriate to slow learners, and it helped in building competence efficiently. Besides, it was easy to seek clarifications from the trainer, and this helped them gain new knowledge every day. However, there are time constraints since one trainer could handle only one employee at a time. It was, therefore, difficult to capture and share knowledge with multiple persons. As a result, there was time-consuming, which affected the delivery of projects. This technique was not efficient, and it required people to handle numerous employees. According to Azam, Khilji, and Khan [5], the method is only applicable when teaching is occurring at the network in the site, and participants implement the knowledge acquired immediately. The Swedish Security Service notes that Swedish police apply the OJT method in their training, but this method goes together with “hands-on” works, which include self-defense. However, the problem is that this method requires retraining the employees after six months. Demirsoy and Petersen [7], note that the OJT

method should be combined with question-based sessions (QABS) and involve new employees by teaching them the guiding principles of the organization. QABS should help employees obtain knowledge, and they would also learn from each other because they belong to the same level. Because OJT may not incorporate multiple employees, QABS will resolve the problem since all employees will build each other, which lays a strong foundation for relations among the employees.

2. Plisse's KM tool poses communication problems between the employees and their managers. Sharing knowledge at the company depends on various success factors of communication and technology. Although technology has significant impacts on the sharing and distribution of resources, communication is fundamental, especially in such a company that involves many departments. The KM tool is entirely digital as employees' access and communicate with one another online. It is the central system of the organization with interconnected components where employees can interact, chat, and discuss various matters online. The problem with this system is that it eliminates the face-to-face interaction between the project managers and the subordinates. The lack of face-to-face hinders communication, especially when physical presence is required to address some issues affecting performance and production [3]. At Plisse, communication is vital, and every employee should participate in the running and operations of the failure. Digital communication is faster, but it weakens relationships among employees. For a company to achieve its objectives, it should provide a way through which its employees understand one another and create friendship for the benefit of the organization. Considering knowledge management at KPMG (Klynveld Peat Marwick Goerdeler), one of the success factors was the creation of a Knowledge Transfer (KT) team that heading the sharing of knowledge and implementation of the new skills acquired [5]. We may not achieve such results at Plisse because of the digital aspect. Therefore, the company should establish a team in charge of the development of KM to have successful knowledge management at Plisse. The team will build high quality and robust management to ensure that all employees access the right knowledge at the right time.
3. The sharing method at Plisse also has issues with the project team structure. The project team has adverse impacts on the sharing of knowledge within the human resources department of the organization. Projects at Plisse are internal more connected with one another, and since every employee has specific expertise, then it becomes hard to share information among the employees. According to Feljan et al. [6], the knowledge management model should provide solutions rather than create problems within the organization. One way in which the project manager collaborates with his employees is through a scrum master, which applies the method of agile program. The purpose of such a model should be to reduce deployment time, operation costs, and enhance the adaptability of the employees in any new environment. Feljan et al. [6] further add that a useful KM model should combine both declarative and procedural knowledge and provide better means of interaction between project teams. The general idea of

knowledge management should give every employee a chance to contribute their ideas to enhance project delivery. However, if there is a model that limits its employees, then it should be eliminated because it will delay the project and lower the performance and efficiency of the employees. An efficient model should connect competencies depending on the existing and utilize the available resources to develop the organization.

4. Plisse is highly dependent on its web portal because they are web-driven, and it is an essential component of competence development. It is one of the effective methods of building competencies among groups, and it tries to generate standards that can be applied by all employees in the organization because it outlines the structure of the organization and how various departments operate. However, this method leads to a lack of motivation towards the employees [4]. One of the roles of the human resources department is to boost the morale of the employees to increase productivity. This technique fails to integrate all elements of the organization, which is critical to the success of Plisse. As a result, the employees do not benefit from technology as expected. According to Azam, Khilji, and Khan [5], any KM model should motivate the employees and improve their decisions. Taking an example of Apple Inc, the KM model is employee-oriented, and it is designed to support the business structure and management of the employees. Apple utilizes four phases, namely infrastructure, KM system, Design, and Development, Deployment, and the last stage is Evaluation. These phases are critical to the performance and interaction of the employees within the organization. If Plisse applies the same concept in its KM tools, then it will motivate its employees, which will be manifested in the production.

3.2 Summary

This part discussed the main challenges identified in the application of the KM. This paper will emphasize on other components that Plisse should consider to improve the utilization and increase benefits from the web-based systems.

4 Recommendations

In this part, we will discuss possible solutions by referencing the literature of the research papers to improve knowledge management at Plisse. From there, this paper will suggest a plan for the human resources department to strengthen cohesion among the employees and improve their performance and efficiency.

4.1 Main Consideration for Knowledge Management Systems

The main idea is that some of the web-based tools failed to consider factors of communication, derailment of the project teams, time consumption, lack of motivation of the tools used. The recommendations for the knowledge management plans are categorized into three main areas:

1. Knowledge sharing considerations
2. Work planning considerations
3. Innovations' considerations

Knowledge Sharing Considerations

- Knowledge management in Plisse is essential for the general development and growth of the organization. Every department has its way of accessing and storing the information received from other sections of the organization. However, the process of knowledge is yet another significant aspect as far as Plisse is concerned. Seemingly, web-based methods such as Plisse's web portal are more concerned with access to expertise, but there are no clear ways through which employees in the human resources department can share knowledge.
- The human resources department should establish measures of not just accessing the information but making the knowledge beneficial to the employees [6]. This approach will guarantee employees learn from the information provided and apply it in their various assignments within the department.
- Plisse should address the limitation observed concerning the spread of knowledge by focusing on the time of delivery of projects. After completing the projects, all members should send their views on how to improve on the next project through the human resources manager. According to Arif and Rahman [10], job satisfaction companies and organizations is very critical, and this can be achieved only through an employee-oriented knowledge management model.
- According to Gharamah, Noordin, Ali, and Brohi [11], some organizations do not understand the importance of knowledge management unless they have a manager. In this case, the manager will convene a meeting where everyone will attend and highlight the essential aspects of significant lessons learned from the previous projects. According to Ngoc-Tan and Gregar [13], this practice will eliminate the time taken by the employees to go online and learn on their own.
- Knowledge sharing is more effective when employees are together face-to-face rather than chatting online and sharing information. One of the challenges of the system used at Plisse is that it eliminates physical interaction among the employees. Plisse can solve this problem by conducting weekly meetings as the human resources manager, where all managers meet with the employees. This move will help in creating a culture of sharing knowledge and creating policies favorable to the organization [10]. According to Feljan et al. [6], the personal

interaction between leaders and their followers is significant because it will also serve as motivation and inspiration to the employees. Such meetings will allow managers to learn from their subordinates who are the ones who usually perform the actual projects, and they witness firsthand challenges.

Work Planning Considerations

- The sole purpose of any knowledge sharing tool should be to improve workflow in any organization and enhance performance and hence, the productivity of the job. An efficient model should increase the accessibility of information that enables employees to learn more about the organization and what the manager expected them to do [5]. It means that information should be useful to them not for personal use but for developing the organization.
- Planning at Plisse is a fundamental aspect considering the high number of employees that the human resources must deal with, and meet their daily needs, with the aim of improving the welfare of the employees [8]. Therefore, the web-based tool such as wikis should be redesigned such that managers from various departments can share their opinions on how these employees can plan their work more effectively. Azam, Khilji, and Khan [5] notes that in Siemens, knowledge management focuses on the engagement of all employees, which changes the way of working. To achieve this, human resources should introduce a PLA model, which stands for Planning, Leading, and Adapting.
- PLA is a method used in Apple to manage the knowledge while at the same time guiding them into becoming more productive. Human resources should establish robust policies and logic to collect information from their customers and shareholders. By incorporating customers into the model, the employees will be working towards meeting their demands and needs.
- Besides, every employee should be a leader in the organization. Demirsoy and Petersen [7] claim that an organization should not restrict anybody in sharing any information with other people, which is deemed necessary for the growth of the organization.
- Ensure there is continuous improvement in the human resources department by ensuring that employees read more and conduct research about their work, which will foster a culture of growth in the department.

Innovations Considerations

- Plisse should emphasize innovation more into its organization, primarily due to the changing nature of human resources. The world is becoming more competitive as companies find new ways of managing their human power and improve the welfare of its people.
- The culture of innovation is essential in the management and sharing of information, and it is the best approach to supporting and implementing more

efficient tools of knowledge management. Azam, Khilji, and Khan [5] note that innovation is a fundamental culture in the organization.

- Plisse should inspire its employees to establish new ways of retrieving information that would be helpful to the report. This includes visiting other organizations and understanding how they best utilize the information available. Azam, Khilji, and Khan [5] notes that in Telenor Telecom Software, there is a database where employees post their views from other companies, which helps in the general development of the company. They can also read about management styles in prominent companies around the world and come up with related or better ways of handling employees [11].
- The human resources may give every employee in their departments a chance to become a group leader in their respective sections. Arif and Rahman [10] suggest that they would later compile reports about the abilities of their colleagues, as they view them from the supervisory angle and suggest how every person should improve the operations of the organization.

4.2 Recommended Implementation Plan

This plan comes from the research findings, and it highlights areas that the human resources department should focus on to improve its function of ensuring the welfare of the employees by enhancing knowledge management. The human resources department is advised to consider the following aspects:

- Reevaluate the operations and activities of the human resources department and its journey in implementing the knowledge management tools.
- The department should incorporate more and share knowledge. This approach includes expanding the source of knowledge and takes views from all stakeholders, including the suppliers and the community.
- Take feedback from the employees on the best ways of managing knowledge.
- Draft a plan which incorporates all managers in the organizations on how they can share views concerning knowledge management at the human resources department.
- Formulate a framework that considers the three significant factors of consideration to improve knowledge management, as discussed in the previous section.
- The department must make long term planning about its employees; in specific, how it is going to improve its operations.
- Self-assessment should be conducted every three months to evaluate the knowledge of the employees.

The above steps lay a platform over which the human resources can best utilize knowledge shared from other departments. Besides, it can use its web portal to continue the interaction among the employees, which strengthens the bond among

the employees. Moreover, it can consider the following actions to eliminate any barrier in its operations:

- Integrate all activities and commitments to exploit the core values of the company, gain competitive advantage, and maximize its value.
- Emphasis research and development to continue delivering quality services to its customers.
- Enhancement of its products in all areas, including production, marketing, and sales.

4.3 Recommended for the Overall Organization

With reference to the primary and secondary research question identified earlier, knowledge management in Plisse is fundamental in its survival of the highly competitive industry of fashion design. The human resources manager, together with the senior managers, should establish plans of considering the three major areas identified in this paper. The burden should not be left to the human resources department alone, but it is the responsibility of the entire company. The manager should focus on the critical factors when sharing information within Plisse, as this is the only way of ensuring optimality and maximum production of Plisse.

4.4 Summary

In the previous part, we discussed areas that the organization should focus on improving knowledge management systems as applied in other successful organizations. The paper categorizes the recommendations into three main parts, including knowledge sharing perspective, work planning aspect, and innovation perspective. After that is an overview of the implementation plan, which presents how the human resources department can utilize and expand its knowledge by incorporating other departments. At the end of the paper, there is a general recommendation that can be used by the whole company. Finally, it is worth mentioning that this study and its produced outcomes and recommendations have been shared with Plisse's decision-makers and concerned parties.

References

1. Salloum SA, Al-Emran M, Shaalan K (2018) The impact of knowledge sharing on information systems: a review. In: International conference on knowledge management in organizations. Springer, pp 94–106

2. Al Mansoori S, Salloum SA (2021) The impact of artificial intelligence and information technologies on the efficiency of knowledge management at modern organizations: a systematic review. In: Al-Emran M, Shaalan K, Hassanien A (eds) Recent advances in intelligent systems and smart applications. Studies in systems, decision and control, vol 295. Springer, Cham
3. Almansoori A, AlShamsi M, Salloum SA, Shaalan K (2021) Critical review of knowledge management in healthcare, pp 99–119. https://doi.org/10.1007/978-3-030-47411-9_6
4. SK Areed, Salloum SA (2021) The role of knowledge management processes for enhancing and supporting innovative organizations: a systematic review. In: Al-Emran M, Shaalan K, Hassanien A (eds) Recent advances in intelligent systems and smart applications. Studies in systems, decision and control, vol 295. Springer, Cham
5. Azam M, Khilji AB, Khan W, Commerce MS (2016) Knowledge management as a strategy & competitive advantage: a strong influence to success (a survey of knowledge management case studies of different organizations). Public Policy Adm Res 6(10):1–15
6. Feljan AV, Karapantelakis A, Mokrushin L, Liang H, Inam R, Fersman E, Azevedo CRB, Raizer K, Souza RS (2017) A framework for knowledge management and automated reasoning applied on intelligent transport systems. ArXiv Preprint ArXiv: 1701.03000
7. Demirsoy A, Petersen K (2018) Semantic knowledge management system to support software engineers: implementation and static evaluation through interviews at ericsson. E-Informatica Softw Eng J 12(1)
8. Poór J, Juhász T, Machová R, Bencsik A, Bilan S (2018) Knowledge management in human resource management: foreign-owned subsidiaries' practices in four CEE countries. J Int Stud 11(3)
9. Zbuchea A, Pînzaru F, Busu M, Stan S-O, Bârgăoanu A (2019) Sustainable knowledge management and its impact on the performances of biotechnology organizations. Sustainability 11(2):359
10. Arif FNT, Rahman SA (2018) Knowledge management and job satisfaction. Int J Acad Res Bus Soc Sci 8(9):266–274
11. Gharamah A, Noordin M, Ali NI, Brohi IA (2018) Knowledge management practice in private sector: building the way for saudi arabia strategic growth and transformation to knowledge-based economy. Int J Eng Technol 7(2)
12. Abualoush S, Masa'deh R, Bataineh K, Alrowwad A (2018) The role of knowledge management process and intellectual capital as intermediary variables between knowledge management infrastructure and organization performance. Interdisc J Inf Knowl Manage 13:279–309
13. Ngoc-Tan N, Gregar A (2018) Impacts of knowledge management on innovations in higher education institutions: an empirical evidence from Vietnam. Economics and Sociology
14. Torabi F, El-Den J (2017) The impact of knowledge management on organizational productivity: a case study on Koosar Bank of Iran. Procedia Comput Sci 124:300–310
15. Raudeliūnienė J, Davidavičienė V, Jakubavičius A (2018) Knowledge management process model. Entrepreneurship Sustain Issues 5(3):542–554

Applying Knowledge Map System for Sharing Knowledge in an Organization

Fatema Hamad Mezahem, Said A. Salloum, and Khaled Shaalan

Abstract Knowledge management (KM) is one of the most important systems that need to be applied in an organization. It has benefits for improving efficiency and productivity in a company to achieve organizational goals. One of the tools that will make finding knowledge easier for employees is the knowledge map. It allows for visualizing the knowledge and link them by creating relationships between domains. The selected organization has several branches in different locations where it hard for the employees to reach the knowledge. The studied organization has applied knowledge management practices however there are issues with the current system. The four main problems related to KM are less communication with decision-maker, no standards for applying KM, not all staff apply KM practices, no real process of converting tacit to explicit. The suggested solutions are to apply the SECI model which is the act of socialization, externalization, combination, and internalization, and implement knowledge mapping. The paper will provide steps for implementing the suggested solutions.

Keywords Knowledge map · Knowledge map and business · Knowledge map and organizations · Knowledge management · Knowledge management systems · Knowledge sharing · Knowledge implementation

F. H. Mezahem · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_80

1 Introduction

1.1 Organization Overview

The studied organization is a governmental organization which for confidentiality reasons the name will not be provided. Like other organizations, their goal is to become one of the Top leading organizations in the UAE. The main task for the organization is to provide auditing services to other organizations in order to prevent deception and corruption. It was established in 1976. The main goal for auditing is to ensure honesty and integrity in other governmental or semi-governmental organizations. There needs to be a social responsibility to perform this task. Auditing is not the only thing that makes the organization. Other operational level departments are located in different braces of the organization. As the main task for the organization is auditing it works with 70 organizations to audit which are mainly government or semi-government companies. The organization gains its authority from the UAE law which gives them power over other organizations. The organization aims to grow and work with other companies such as the private sector. The main goal is to protect the county from corruption and increase the performance of the organizations in the UAE. The organization generates reports that include audit results and the amount of reliability of materials delivered by organizations. This will help decision-makers in the country to make good financial decisions. The organization has several branches located in different cities. Most auditors work offsite and few work in their offices. Communication between branches and employees needs to be maintained. This leads to having issues with Knowledge management where there needs to be a system implemented KM practices. Figure 1 below simulates the organizational chart of the studied organization.

1.2 Knowledge Management in the Organization

The selected organization can beneficial from KM to improve the business process. Employees are required to work outside their office which makes them had to

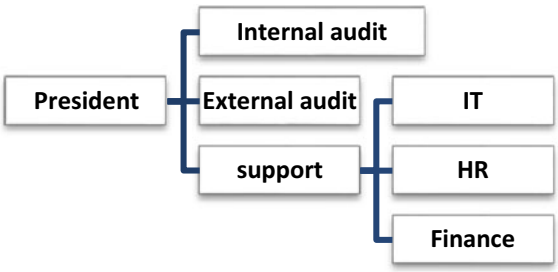


Fig. 1 Organizational chart

communicate with their managers. Also, employees need to communicate with each other to share knowledge. Some knowledge-sharing practices have been applied in the organization. Therefore, experts are encouraged to share knowledge with other employees. The knowledge is shared using meetings training, face-to-face interaction, emails, and documents. The organization has several departments where each department has confidential information that needs to be secured. The case study focuses on knowledge sharing between each department such as the IT department, HR, finance, etc. Some employees are more knowledgeable than others due to their different experiences. Other employees have a special task or a system where they have been working on it for years. If these types of employees leave the organization whether they retire or found a different job, their knowledge will be lost with them and no one will know how to run the system. In the organization, basic knowledge transfer has been applied where the expert employee trains another employee and confirms the knowledge transfer by email. However, using this simple method does not guarantee whether the knowledge has been fully transferred to the next employee. It depends on the experts if they are willing to share their knowledge. Some might not be interested in sharing or they believe it not their responsibility and it is a time-consuming task.

2 Related Work

2.1 Characteristic of Knowledge

According to Lin [1], knowledge has characteristics of spreading, developing, reusability, and guidance. To manage the knowledge they need to be cooperative between the employees in the organization. In some cases, it might be useful to use the help of experts from inside and outside the organization. Knowledge is continually changing depending on different factors such as rapidly changing technology, new researches, and social communication, perspectives, and social interactions [2]. Organizations can benefit from the knowledge they gained from previous cases to solve similar and new cases. The experience from the expert employees can be captured and share among employees which will speed the organizational process.

2.2 Knowledge Management

With the rise of the global economy, organizations are now seeking a way to compete in the global market [3–9]. This means that organizations need to have a competitive advantage and improve the organizational process. Knowledge management provides an effective approval to benefit from the knowledge and improve

an organization's capability [10–13]. The business environment is changing rapidly and the organization should accept these changes and find solutions for new problems using knowledge management tools. The internet provided an advantage for acquiring knowledge and sharing them over a global distance [14]. In organizations, most of the work is accomplished by teams where they share the experience. Most of the knowledge is in a tacit form inside a person's mind. The role of knowledge management inside an organization is to identify this knowledge and transform it into a form that can be managed. Digitalizing the knowledge is also useful which makes it easier to share and manage by individuals in an organization [15]. Many organizations implemented knowledge management as part of their business process. Knowledge is continually formed within the situation of any company's processes [16]. Large organizations acquire a great number of knowledge resources which made it difficult for employees to find the resources and get the knowledge. As a result, knowledge users need to spend a lot of time searching for the knowledge they need and share it with the right people [14]. There should be a mechanism for organizing knowledge to make it easier for employees to find their resources of knowledge. One of the best solutions for organizing knowledge is the use of knowledge maps.

2.3 Knowledge Map

A knowledge map is a diagram that has the suggested actions an employee may apply in a case and the knowledge they proceeded when they applied the solution [17]. Knowledge in a knowledge map represents the relationship between causes and effects. Knowledge maps represent knowledge in a visual format which makes it easier to analyze and retrieve knowledge [18]. Many researchers view knowledge maps as one of the main components of knowledge management. A knowledge map is considered a graphical architecture in the knowledge field. Visualization is one of the characteristics of knowledge maps. Using this approach in knowledge management will simplify the process and make it more user-friendly [1]. Knowledge mapping allows for easy human and machine interpretation and can be implemented on complex tasks. According to Kang et al. [18] it will enable the employees to examine the knowledge on a universal measure and using different perceptions. In addition, knowledge maps are used to find relationships between knowledge which makes it easier to retrieve. Knowledge maps can help knowledge users, decision-makers, and developers. It will assist decision-makers to analyze the standing of the knowledge and plan for future knowledge progress. Different strategies can be used for knowledge development. There is no single form of knowledge maps that fits all organizations however it can be utilized according to a specific domain [15]. It can show the path for action to achieve a specific goal. One is holistic testing, mapping tools filter knowledge using different factors and classifying knowledge. Knowledge map improves knowledge management in areas such as building, learning, and distribution [19]. One of the tools used for

knowledge management is the knowledge map. According to Liu et al. [14], the current knowledge maps are visualizing user-friendly format which uses software programs that are combined with data mining processes. Knowledge mapping is a challenging task that needs organization, and classification techniques. The task needs to identify the most important organizational knowledge and use it as a base for classifying other knowledge [14]. Based on the type of the organization the number of groups varies with the type of resource. In addition, the type of items in the map varies such as text, images, audios, videos, and database. The goal of knowledge map tools is to find relationships between these knowledge resources. This can be done by linking topics, themes, and ideas for each type of item [14]. Hence, the effective mapping module can be conceded as an added knowledge for the organization.

2.4 Knowledge Map Classification

Classification and clustering are the major techniques for knowledge mapping. Many researchers believe that automatic clustering and classification methods for knowledge mapping are efficient and can be applied to organizations [15]. Heimerl et al. [20] developed as a system called Word Cloud Explore, that uses linguistic knowledge which will analyze a text or a document and link with topics and find relationships with words. However, the method has its problem which is that it is concentrating on linguistic type information and less usage of other types such as images and videos. The classification for knowledge map is done by Applying machine-learning algorithms to group documents. According to Liu et al. [14] VSM is a technique used to represent documents as a multidimensional vector, each dimension is linked to a unique topic extracted from the source. Another clustering method is the Self-Organizing Map (SOM) which is a type of neural network algorithm that can produce a hierarchical knowledge mapping scheme. Knowledge map applications analyze the links between nodes and context [1]. According to Lin [1] knowledge mapping has two types of nodes: articles and keywords. Many links may occur among two documents because of containing several common keywords. Other documents may not have links because it does not have common keywords [1]. Identification and Evaluation of Alternate Solutions Both research papers that we have investigated tackled the role of having a knowledge management system in a service provider environment and the positive impact it has when operating in a highly competitive and dynamic market. The two case studies that we are tackling in this paper are one of Cordova's that followed a well through knowledge management assessment tool, model, and knowledge management systems to sustain in a competitive market. The second is the Meittenen case study where the service company failed to embed any knowledge management model, mechanisms, or tools.

3 Problem Identification and Analysis

The selected organization applies some of the knowledge management practices however it has many issues which need to be addressed. Addressing the problems will help in finding the best solution and applying best practices in all aspects of the organization. The major problems are less communication with a decision maker, no standards for applying KM, not all staff apply KM practices, no real process of con-verting tacit to explicit.

1. Less communication with the decision-maker
   In order for managers or decision-makers to make the best decision, they need to know the knowledge behind it. The need to be in contact with the experts related to the case. For example, if the case is related to purchasing a new IT system, they will contact experts who know the technology or employees who have used similar systems. In the organization is not easy to know who the experts are or even contact employees who use the system. The reason for this is that there is not a practical KM system applied and communication is hard because they work in a separate branch. As a result, managers will not be able to make the best decision which leads to the loss of money, time, and workforces. The decision-maker usually has a great influence on other employees so, they need to be aware of the KM practices. They need to be the first to apply knowledge sharing and encourage others to share their knowledge. If the knowledge is shredded effectively better decisions and actions will be made by managers which leads to better performance. For example, if the decision-maker knows the technology behind the system they will know if it is worth purchasing or not. When contacting employees who used similar systems, they can get information on how they are using it and what problems they are facing. The goal is to avoid these problems and improve the process. Knowledgeable employees can give suggestions to add or remove unwanted features.
2. No standards for applying KM
   Although the knowledge is shared between individuals to some extent, in the studied organization, the is no standard way for sharing the knowledge. Some knowledge is shared verbally and others are documented. It is no easy way for employees to find the knowledge when they need it. They don't know if the knowledge is available in the organization or not. They don't know who to reach out to in finding a solution for a problem. This issue is seen in the help desk in the IT department where employees are not able to benefit from the solutions of older and similar issues that they face. They act with all cases as new cases a search for solutions individually. There needs to be a system that allows them to look at older cases and add new solved ones to a specific KM system. Another issue that some knowledge sharing happens between individuals and other team members who cannot benefit from the knowledge because it is not documented or recorded. The knowledge needs to be in a demilitarized form to make it possible to manage and share across the organization or domain. Some tacit knowledge cannot be transformed to explicit and in this case, employees should

be able to know how to reach for this person to retrieve the knowledge when they need it. This requires that individuals in the organization have good communication skills.

3. Not all staff apply KM practices
   Some employees do not share their knowledge in the organization for many reasons. First, that they are not aware of the importance of the KM process. The employees believe that it is a time-consuming task and they will not benefit from it. The other problem is that they don't want to share their knowledge. They think that sharing the knowledge will minimize their role in the organization [21]. For some reason, these types of employees don't trust the people around them. They do not interact with other team members and prefer to work individually. The final issue is that employees do not know how to share the knowledge they have or explain it to others. They lack the skill that educating and understandably present their knowledge.
4. No real process of converting tacit to explicit
   Tacit knowledge is described as personal knowledge which is resulting from the person's activity and experience [22]. Most of the knowledge transfer processes happen in a face-to-face meeting and no real documentation is happening. There is no process for knowledge searcher in the organization to search for knowledge. In other words, they don't know where to search or who to ask to find a solution for a specific problem. The current cycle in the studied organization is that an expert shares some of their knowledge with other employees and it is still in a tacit form. Therefore, the risk of losing the knowledge from one or both of the knowledgeable employees is possible. Employees can leave the organization whether they retire or find a new job and, their knowledge will be lost with them. The organization should apply an effective Knowledge management system to capture the knowledge and store it in a way that all employees can benefit from it. They can use the SECI Model. One of the solutions is the use of knowledge maps which links knowledge in a graphical form which makes it easy to use and findinformation.

4 Identification and Evaluation of Alternate Solutions

The solution for the above mention problems is to apply a knowledge management system in the organization. The organization can benefit from the SECI Model in socialization, externalization, combination, and internalization [23]. Part of the knowledge management system is to use knowledge maps. Knowledge maps allow employees to find the source of knowledge they need easily.

1. Apply KM organization culture
   Another solution is to apply a knowledge management culture. The culture will make knowledge sharing part of the organizational process. In this culture, all employees are required to share their knowledge as it is expected from them.

The employees who used to refuse knowledge sharing will change their attitude according to the sounding culture. Decision-makers should the first to apply knowledge sharing which results in encouraging another follower to do the same. The need to make employees comfortable to share their knowledge and ensure that the KM system is here to help them and not replace them. By doing this trust among employees will increase and teams will produce more efficient work.

2. Aware employees of the benefits
   When all employees share their knowledge, everyone will benefit as each individual will gain more knowledge and, increase their efficiency in the organization. When employees find themselves improving, they will be satisfied and will be more productive the organization improve and achieve their goals faster.
3. Use KM tool
   Applying an effective KM system is the key to solving-knowledge related issues. In the studied organization the idea of knowledge sharing is not new however it lacks a convenient method. In this organization knowledge sharing should be in two levels organizational lever where all employees can view the knowledge and department level where specific teams can share confidential knowledge. The nature of the organization requires confidentiality as it has sensitive governmental information. For example, the IT department should not be able to access knowledge related to auditing tasks or the other way around. On the other hand, they should know where to go in case they need another type of knowledge even if not related to their domains. For employees to access the knowledge, they will have to get approvals from top management. Another use is in the case of managers where they need to know which expert to reach to in making an action. Using a tool that will organize these tasks will be the best solution for the problem.
4. Knowledge maps to make it easier to find knowledge
   Maps are an effective way to link knowledge by applying the algorithm to find common keywords [1]. Knowledge maps will represent knowledge in a visual format which makes it easier to analyze and retrieve knowledge [18].

5 Implementation Plan

To implement knowledge maps the company needs to purchase a tool that will create maps automatically with the help of little manual work. The manual work will be in adding the knowledge to the system and assigning keywords. This task can be done in all departments by training the existing employees to register their knowledge. Another solution is to assign a specialized employee to capture and collect knowledge from experts in the organization. For extracting knowledge interviews can be used with questionnaires for experienced employees. Not all employees know where to find the knowledge they need. A knowledge map is a

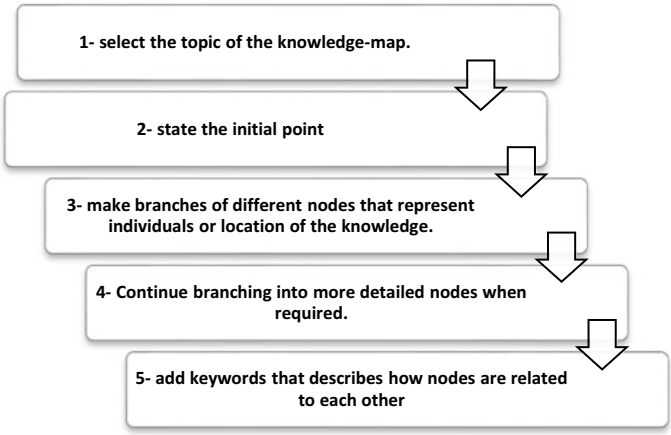


Fig. 2 Steps for creating a knowledge map in MindManager

perfect tool for finding people in an organization who have a specific experience in a subject [18]. Knowledge mapping will help employees to find the knowledge they need and from the right people. One of the efficient tools for creating knowledge maps is MindManager [24]. It is one of the simplest and easy use tools. MindManager has installed templates the will helped in creating a quick and easy knowledge map. In addition, it allows to creation customized knowledge map. The steps for creating a new customized map is simple and is demonstrated in Fig. 2

To draw the knowledge map, can include the number of domains, topics, knowledge resources, and the relationship between them [14]. It is challenging to illustrate all types of knowledge and all links between them in one map. To cover all aspects of knowledge in the organization in detail. We need to state domain as the initial component to establish knowledge and its relationships. The domain is where the knowledge is used or, where an action, process, or business is happening. It will be more useful to create domain-based knowledge mapping. The topic in the knowledge map describes the knowledge source by its occurrences. Topics in domains are connected by links. According to Liu et al. [14] there are three methods for creating knowledge maps which are individual knowledge tagging, domain topic selection, inter-domain topic association analysis. First, is individual knowledge tagging: this procedure is to create a personalized knowledge map and to shape knowledge of interest. The individual who works I a specific domain creates the document and assigns labels that define the content of the document. Next, is the domain topic selection where a domain will show the content of knowledge items and their relationship a domain appears when a seitan action or process is established? Also, the tagged item will contain several exclusive terms, words, or expressions that are tagged on documents. Finally, inter-domain topic association analysis in this process it will search for similarity which will relate the topic of several domains.

6 Recommendations

Knowledge mapping fulfills most of the business needs for the organization. Some more manual effort will benefit the organization in perfecting the model. Knowledge maps can be a tool used by all employees and decision-makers to benefit from the knowledge. The system will benefit the organization in improving cost, effectiveness, and efficiency. The success of the knowledge maps system depends on the individuals who apply it [25]. However, the method has its own problem which is that it is concentrating on linguistic type information and less usage of other types such as images and videos. For example, less time will be needed to find the knowledge and when having access to the required knowledge actions and decisions will be improved. This will result in better management of recourses and cost. Employees who work in the same department usually have similar goals and can benefit from each other experience. The experience should be shared and linked to the knowledge map to make it easier for other employees to find it. The knowledge map presented enables us to understand the structure of the information and knowledge in the studied organization and find relationships among them. Knowledge maps can be applied at the organizational level or department level.

7 Conclusion

In conclusion, it is necessary for organizations nowadays to implement knowledge management systems to compete in the global business. Knowledge management will increase the productivity of the organization and will result in more efficiency when dealing with recourses. The paper provided solutions for different knowledge management problems in a targeted organization. One of the main solutions is to apply knowledge management maps. Using maps will enable to visualize knowledge and find relationships between domains and document using keywords. Employees will be able to find the knowledge they need easily and share their own. Creating the knowledge-sharing environment is one of the initial tasks for the organization to ensure employee satisfaction and encourage knowledge-sharing behavior.

Acknowledgements This work is a part of a project submitted to The British University in Dubai.

References

1. Lin FT (2019) Drawing a knowledge map of smart city knowledge in academia. Knowl Organ 46(6):419–438
2. Haraldsdottir RK, Gunnlaugsdottir J, Hvannberg ET, Christensen PH (2018) Registration, access and use of personal knowledge in organizations. Int J Inf Manage 40:8–16
3. Salloum SA, Al-Emran M, Shaalan K (2018) The impact of knowledge sharing on information systems: a review. In: 13th International conference, KMO 2018

4. Al Mansoori S, Salloum SA (2021) The impact of artificial intelligence and information technologies on the efficiency of knowledge management at modern organizations: a systematic review. In: Al-Emran M, Shaalan K, Hassanien A (eds) Recent advances in intelligent systems and smart applications, vol 295. Springer, Cham, 2021
5. Areed S, Salloum SA (2021) The role of knowledge management processes for enhancing and supporting innovative organizations: a systematic review. In: Al-Emran M, Shaalan K, Hassanien A (eds) Recent advances in intelligent systems and smart applications, vol 295. Springer, Cham
6. Almansoori A, AlShamsi M, Salloum SA, Shaalan K (2021) Critical review of knowledge management in healthcare, pp 99–119
7. Mehrez AAA, Alshurideh M, Kurdi BA, Salloum SA (2021) Internal factors affect knowledge management and firm performance: a systematic review, AISC 2021, vol 1261
8. Alsharhan A, Salloum S, Shaalan K. The impact of e learning as a knowledge management tool in organizational performance
9. AlGhanem H, Shanaa M, Salloum S, Shaalan K (2020) The role of KM in enhancing AI algorithms and systems. Adv Sci Technol Eng Syst J 5(4):388–396
10. Arpaci I, Al-Emran M, Al-Sharafi MA (2020) The impact of knowledge management practices on the acceptance of massive open online courses (MOOCs) by engineering students: a cross-cultural comparison. Tele Inf
11. Al-Sharafi MA, Herzallah FAT, Alajmi Q, Mukhtar M, Arshah RA, Eleyan D (2020) Information culture effect on e-commerce adoption in small and medium enterprises: a structural equation modelling approach. Int J Bus Inf Syst 35(4):415–438
12. Al-Emran M, Mezhuyev V, Kamaludin A (2020) Towards a conceptual model for examining the impact of knowledge management factors on mobile learning acceptance. Technol Soc
13. Al-Emran M, Teo T (2019) Do knowledge acquisition and knowledge sharing really affect e-learning adoption? An empirical study. Educ Inf Technol
14. Liu C, Li L, Lv J (2009) Social sustainability: towards some definitions. Hawke Res Inst Work Pap Ser 153(27):1–31
15. Wang S (2002) Knowledge maps for managing web-based business. Ind Manag Data Syst 102(7):357–364
16. Zablith F, Faraj S, Azad B (2016) Organizational knowledge generation: lessons from online communities. Bus Process Manag J
17. Yoo K, Suh E, Kim KY (2007) Knowledge flow-based business process redesign: applying a knowledge map to redesign a business process. J Knowl Manag 11(3):104–125
18. Kang I, Park Y, Kim Y (2003) A framework for designing a workflow-based knowledge map. Bus Process Manag J 9(3):281–294
19. Zins C (2007) Knowledge map of information science. J Am Soc Inf Sci Technol 58(4):526–535
20. Heimerl F, Lohmann S, Lange S, Ertl T (2014) Word cloud explorer: text analytics based on word clouds. In: 2014 47th Hawaii international conference on system sciences, pp 1833–1842
21. Wu W-L, Lee Y-C (2016) How to make a knowledge-sharing group: a group social capital perspective. Pers Rev
22. Chen H, Baptista Nunes M, Ragsdell G, An X (2018) Extrinsic and intrinsic motivation for experience grounded tacit knowledge sharing in Chinese software organisations. J Knowl Manag 22(2):478–498
23. Oyemomi O, Liu S, Neaga I, Alkhuraiji A (2016) How knowledge sharing and business process contribute to organizational performance: using the fsQCA approach. J Bus Res 69 (11):5222–5227
24. MindManager (2020) What is a knowledge map?
25. Ebener S et al (2006) Knowledge mapping as a technique to support knowledge translation. Bull World Health Organ 84:636–642

Use of Multi Agent Knowledge Management System in Technology Service Providers

Fatma Ahmad Bazargan, Said A. Salloum, and Khaled Shaalan

Abstract The objective of this paper is to propose the use of a multi agent approach as a knowledge management system in a technology service provider (TSP) environment. The TSP shall ensure that the services provided are reliable and trustworthy to sustain in a competitive technology market. This is achieved by possessing a fleet of expert intellectual professionals who can diligently provide the required services to the clients. Hence, the retention of intellectual capital is crucial to succeed and sustain. Labor mobility presents a problem for TSPs who depend mainly on knowledge workers to provide its core business services. As when these knowledge-based resources leave their jobs they take the knowledge of organizational intellectual work with them. Hence, it is crucial for these TSPs to deploy necessary mechanisms to preserve and manage the organizational knowledge and knowhow to address the labor mobility problem. This study will detail out the solution and the implementation plan to address the labor mobility problem.

Keywords Service companies · Knowledge workers (KW) · SECI model · Tacit knowledge (TK) · Explicit knowledge (EK) · Knowledge management system (KMS)

F. A. Bazargan · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_81

1 Introduction

The current business landscape is highly competitive, dynamic, turbulent, uncertain and rapidly changing. In [1] quotes in an economy where the only certainty is uncertainty, the one sure source of lasting competitive advantage is knowledge. Hence, it is evident that today's economy runs on knowledge and to capitalize on that fact, organizations need to work assiduously towards knowledge management. Knowledge has become one of the most important intangible assets for the organizations [2–7] that strives to maintain a competitive position over time, and provides a sustainable competitive advantage in a competitive and dynamic economy [8–15]. In [16] defines knowledge as a combination of experiences, values, contextualized information and ideas from experts that provide a framework to evaluate and incorporate new experiences and information. Knowledge is also the accrual of insights, ideas, learning, understanding, memory, visuals, sight, cognitive, technical skills, and so on. On the other hand, Liberona and Ruiz [17] argues that only by possessing the knowledge, no matter how valuable it may be, does not ensure the achievement of organizational competitive advantage; knowledge must be developed and managed to produce actions that eventually create value. Hence, knowledge although present in people minds (i.e. tacit knowledge), it also exists within an organization's context. Knowledge can be found not only in organization's explicit knowledge database such as its documentation, manuals, strategies, framework, etc. but also in its processes, practices, routines, and knowhow. Based on characteristic, there are two types of knowledge that is tacit and explicit [1]. Nonaka and Von Krogh [18] defines tacit knowledge as personal, resides in the minds of people and is rooted in an individual's experience and values. It is context-specific, difficult to measure and therefore hard to formalize and communicate. On the other hand, explicit or "codified" or "know-what" knowledge, refers to the knowledge that is objective, rational, transmittable in formal, systematic language, articulated in formal language and frequently takes the form of grammatical statements, mathematical expressions, specifications, documents, reports, manuals, databases, catalogues, presentations, patents, formulas, and so forth [1]. Both forms of knowledge can be articulated to generate decisions and a critical resource for transforming the organizations [19].

Knowledge management (KM) is a process that uses all various forms of knowledge to create new markets, rapidly develop new products and services, respond promptly to customers and dominate emerging technologies. The implementation of KM requires three factors that is organizing the knowledge, the knowledge worker and a mean of information technology [20]. Hence, to achieve knowledge management the three factors need to be achieved. Knowledge management strategy is also crucial to be stated in every organization. According to Hansen et al. [21] there are two approaches to manage knowledge that is codification and personalization strategies. Codification is when the knowledge is codified and stored in knowledge databases, and these databases are accessed by all the employees within the organization. On the other hand, personalization is when the

knowledge is related to the people element and is only shared, transferred and exchanged through direct communication and face-to-face interaction with other individuals. According to Nonaka and Takeuchi [1], knowledge is managed through four different processes that form a cycle. The knowledge is created through the interaction between tacit and explicit knowledge in the organization, this is depicted in Fig. 1.

The four modes of the SECI knowledge spiral are as follows:

1. **Socialization—Tacit to Tacit**: it is the sharing and creating of tacit knowledge directly with another individual. In this mode of knowledge sharing every individual interacts directly with another individual to share and create tacit knowledge. The knowledge is learnt and built upon through observing, imitating, practicing or becoming 'socialized' into a unique way of handling things. This mode of learning is from mentors, peers, professional employees, suppliers, and so on. This can be the least effective knowledge creation mode of the spiral process because the benefit is limited between two individuals or more and no new knowledge has been converted into explicit knowledge for the wider benefit of the organization [1, 2, 22].
2. **Externalization—Tacit to Explicit**: it is the articulation of existing tacit knowledge from human intellect and converting it into explicit knowledge also known as externalization. Thus, allowing this newly documented knowledge to be shared with wider audience of the organization. The knowledge is learnt and built upon through recording discussions, descriptions, innovations, and "know-how" and documenting it into manuals [1, 2, 22].
3. **Combination—Explicit to Explicit**: it is the combining of distinct pieces of newly introduced explicit knowledge into a new whole knowledge within the organization. Both combination and externalization equip an organization to create new emerging markets and introduce new products and services within the organization to gain competitive advantage [1, 2, 22].
4. **Internalization—Explicit to Tacit**: it is the sharing, articulating, reframing and interpreting of newly introduced explicit knowledge and then internalizing it within the organization. Hence, learning and acquiring new tacit knowledge through practice. This newly acquired tacit knowledge can either broaden,

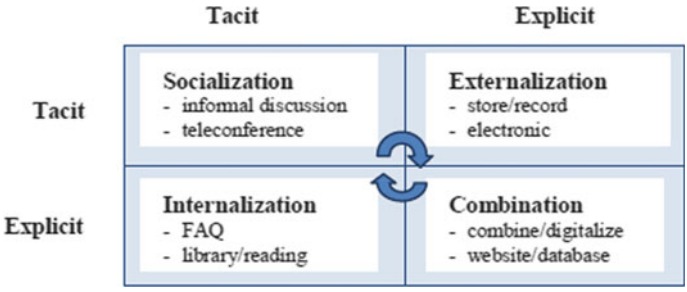


Fig. 1 Knowledge management spiral process [1]

expand, or build on the current existent tacit knowledge of the human intellect [1, 2, 22].

Knowledge management systems (KMS) is an information system developed to facilitate and support the organizational processes of knowledge creation, storage, retrieval, transfer, and application [23]. That is, the main purpose of KMS is to create organizational intellectual knowledge that will eventually create organizational value. According to Alavi and Leidner [23], there are three salient features of KMS: knowledge repositories, knowledge maps, and collaborative tools.

1. According to Davenport and Völpel [24], knowledge repositories are the most common type of KMS. Repositories store the codified knowledge and facilitates knowledge re-use. All organizational intellectual work is captured, generated, stored, organized, searched, retrieved, and used. Figure 2 shows [25] knowledge repository contextual model.
2. According to Jiang and Klein [26], knowledge maps are also known as expert directories, searchable indexes or catalogues of expertise held by individual employees. These catalogues provide means of finding and contacting individuals who hold specialized knowledge, expertise and experiences [27].
3. According to Bernard [28] collaborative tools enable knowledge exchange and sharing amongst knowledge seekers and knowledge providers such as groupware, email, chat, electronics forums, intranets, and conferencing.

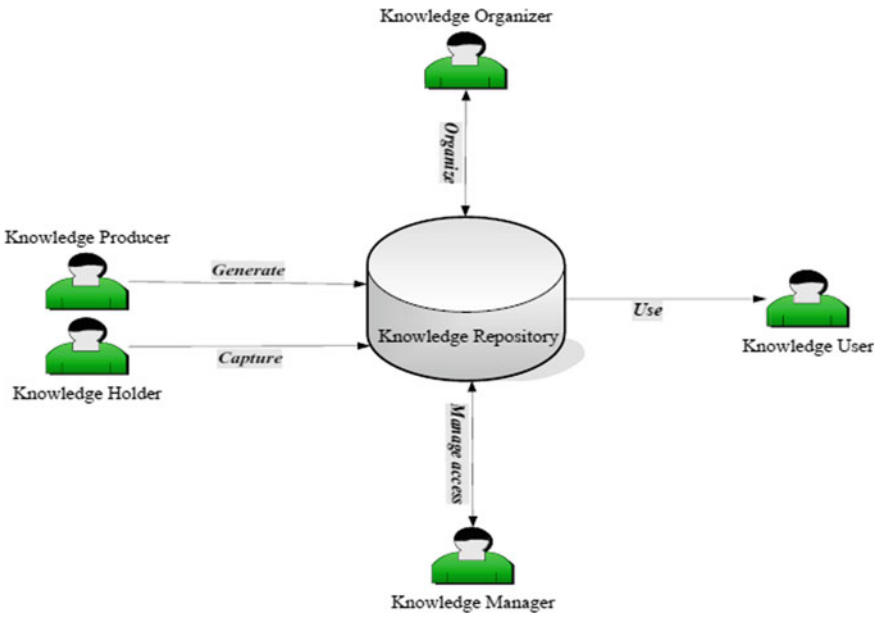


Fig. 2 Contextual model for knowledge repositories [25]

As supported by Nonaka and Toyama [29] organizations differ because they want and strive to differ. These organizations evolve differently because they envision different futures, which are based on their own dreams and ideals, and also because they adopt different strategies and structures to realize such futures. Hence, successful organizations are those that respond rapidly to customers, create new markets, quickly develop new products and services, and dominate emergent technologies. The heart of their success is developing and managing knowledge to produce products and services that creates value which in turn allows them to succeed and sustain in a fierce competitive and dynamic market. The focus of these successful organizations is centered around three subjective elements of organizational leadership vision, its value system and the dedication and commitment of its knowledge workers (i.e. an organization's intellectual capital). In this paper the focus is on the third subjective element of nurturing and preserving the people element that is known as knowledge workers, knowledge-based resources or the organization's intellectual capital. Organizations aim to maximize the value of knowledge as it is important [30]. This is achieved through managing the knowledge that allows the organization to influence the core competences and attain competitive advantage over time. In other words, an organization that develops and manages knowledge effectively throughout the organization has the ability to introduce new products and services that attract new customers whilst being ahead of competition. As organizations move towards developing products and services that are highly dependent on knowledge their reliance on knowledge workers is increased. Hence, the knowledge workers in these organizations become core and considered a valuable asset that needs to be retained and nurtured.

This paper is structured as follow. In Sect. 2, we identify and analyze the problem in the target organization of a technology service provider. The major problem that we are aiming to propose a solution for is presented in Sect. 3. In Sect. 4, we present the solutions that can address the major problems from the research papers that we have tackled. Recommendation of an effective and efficient solution is proposed in Sect. 5 and the implementation plan is detailed out in Sect. 6.

2 Related Work

2.1 Problem Identification and Analysis

Organizations are investing significant amount of time, resources and money annually on information technology (i.e. products, services and systems) towards the development and deployment of state-of-the-art knowledge management systems (KMS). KMS and also known as knowledge-based systems are used to collect, analyze, store, disseminate and retrieve knowledge [31]. In other words, KMS are deployed to facilitate and support the organizational processes of knowledge

creation, storage, retrieval, transfer, and application [23]. However, this massive investment in technology has two major issues. One that the effectiveness of such knowledge management systems depends solely on people to feed in the knowledge, access it and interact with it to create organizational intellectual knowledge. Second, that these KMS only facilitates the storage of one form of knowledge which is explicit. Explicit or "codified" knowledge, refers to the knowledge that is objective, transmittable in formal, systematic language frequently takes the form of documents, reports, manuals, catalogues, presentations, patents, formulas, etc. [1]. However, what organizations neglect is the knowledge that can be captured from their intellectual capital which is its professional employees or known as knowledge workers also known as personalization knowledge strategy. This personal knowledge is known as tacit knowledge. Tacit knowledge is personal, resides in the minds of people and is rooted in an individual's experience and values [18]. It is context-specific, difficult to measure and therefore hard to formalize and communicate [32]. Hence, in order to fully benefit from the collective value of both explicit and tacit, an organization needs to investigate knowledge management systems or approaches that can capture both forms of knowledge the tacit and explicit knowledge. Furthermore, organizations need to comprehend the modes of interaction between these two forms of knowledge in order to effectively develop and manage knowledge.

The target organization in this case study is a local technology service provider (TSP). A TSP is a vendor that delivers and provides information technology solutions and/or services to customers. These customers are considered various local government and private entities from different industry verticals such as oil and gas, energy, healthcare, financial institutions, etc. The IT solutions and services incorporate all information technology businesses that provide products and systems through services that are either on-demand, pay per use or a hybrid delivery model. The various forms of solution and/or services provided by this TSP is as follows:

1. Infrastructure managed services: consisting of network, systems and service desk services;
2. Application managed services: consisting of application management, application enhancement, enterprise monitoring and database management;
3. Cyber defense managed services: consisting of cyber defense applications, cyber security monitoring, cyber security consultancy and advisory services; and
4. Cloud hosting services: various cloud models of private, public and hybrid; and
5. Data center and disaster recovery services offering a tier four data center environment.

The TSP should ensure that the services provided are reliable, robust and trustworthy to sustain in a highly competitive technology market. Moreover, it shall also ensure that the products and services are unique and has a competitive edge to the existing competition. Other core business pillar for the service provider, is that it possesses a fleet of expert intellectual professionals who can assiduously provide

the required services, consultancy and advisory to the clients whenever needed and/or requirement arises. Hence, the retention and engagement of intellectual capital for a service provider is crucial to succeed and sustain in a competitive technology market.

2.2 Major Problems

Service providers are revenue-based organizations those that have a customer base. These customers approach service providers for the niche, quality, reliability, and robustness of products and services that they provide. These service providers highly depend on knowledge workers that is their professional employees to design, develop, provide, and maintain the high class of products and services to sustain in a highly competitive and dynamic market. Hence, continuously retaining, nurturing and engaging these professional employees is crucial for the organization's survival. There are namely two major problems existent in our target organization as follows.

Labor mobility presents a problem for service providers who depends mainly on knowledge workers to provide its core business services. As when these knowledge-based resources leave their jobs they take the knowledge of organizational intellectual work with them. The service provider's pain lies in filling that gap by searching the market, recruiting and training the new apprentices. In many situations, all what they find in the market are novices to fill the positions. Hence, it is crucial for these service providers to deploy necessary solution, processes and mechanisms to safeguard, preserve and manage the organizational knowledge and knowhow to address the labor mobility problem. Another problem pertaining to technology service providers is knowledge workers' engagement within the business fabric of the organization. Service providers highly depend on knowledge workers to create new markets and develop niche products and services. Hence, these knowledge workers need on a continuous basis a live platform and dynamic system to feed in their ideas and suggestions of the next emerging product and service to introduce. Having the professional employees' engagement in the overall success of the organization is crucial. However, lack of having such a platform underwhelms the knowledge workers to produce and be actively engaged. Hence, service providers who need to maintain their competitive position in a highly competitive and dynamic market should strive to address these problems to be ahead of the curve.

3 Identification and Evaluation of Alternate Solutions

Both research papers that we have investigated tackled the role of having knowledge management system in a service provider environment and the positive impact it has when operating in a highly competitive and dynamic market. The two case studies that we are tackling in this paper is one of Cordova that followed a well through knowledge management assessment tool, model and knowledge management systems in order to sustain in a competitive market. The second is the Meittenen case study where the service company failed to embed any knowledge management model, mechanisms or tools.

3.1 First Case Study: Cordova

The first case study emphasized on aligning human capital management with the business strategy and managing knowledge through conversion mechanisms between tacit and explicit knowledge. This emphasis is to translate acquired knowledge from the intellectual capital into information technologies so that the organizational culture is solid and invariable in terms of results where there is employee turnover [33]. The solution used by the first case study of Cordova uses models, methods, and tools for the management of expertise (experience) of the knowledge workers and cooperation and collaboration amongst them for knowledge management. The Cordova case study started by structuring the intellectual capital at the Damage Adjuster company into five levels of expertise which is, novice, advanced beginner, competent, proficient and expert. Then designed a knowledge management assessment tool (KMAT) that has four different factors favoring knowledge management that is leadership, culture, technology, communication and measurement of intellectual capital. The final knowledge management model was one that is known as Dynamic Model of Knowledge Management (DMKM) that is a description of the rotation of knowledge in six various processes of acquire, socialize, structure, integrate, add value and detect opportunities that knowledge provides. Whilst considering people, systems, processes, products and services, environment and the market. The services company KPMG also embraces a similar model for organizational learning and growth where it considers factors of organizational culture, leadership, style, strategy, structure, management of people, and information and communications systems. Another model for knowledge management is that of Nonaka and Takeuchi model that contemplates the ontological dimension. The first dimension is that of generating the knowledge based on the participants in the process. These participants of knowledge function as knowledge creating agents: the individual, the group, the organizational and the inter-organizational level. The second dimension is the integration of the SECI spiral knowledge process of converting and transferring of knowledge between tacit and explicit in term of socialize, internalize, externalize, and combine. These

processes include sharing of values, beliefs, experiences, insights, and expertise through verbal communication, face-to-face, documents, manuals, etc. The last dimension is that of concept of *ba* which is the knowledge that is shared, created and used in order to generate value back as the organizational intellectual work.

The knowledge management tools used in the Cordova case study is through the information technology medium to obtain, process, store, exchange and disseminate information. In order to suite the knowledge management model that was designed and explained previously Cordova used various IT platforms to achieve the results of the model. These tools were the use of:

1. For the purpose of E-learning: video streaming, online forums, curricular nets, and learning management systems (LMS) were used;
2. For the purpose of information repositories and database management systems: knowledge base systems, data warehouses and expert systems were used;
3. For the purpose of corporate intranets: survey systems, content management, user profiling systems, and visits statistics were used
4. For the purpose of collaboration tools: wiki systems, chats, collaborative forums, and mailing systems were used;
5. For the purpose of customer intelligence: CRM, multi-channel care platforms, customer database were used;
6. For the purpose of process management: online manuals, procedural bases, workflows, RSS news feeds were used;
7. For the purpose of intelligent agents: search engines, management reports systems, talent management, and multidimensional cube were used; and
8. For the purpose of collaborative platforms: expert guides, CMSM, project support and ECM were used.

Cordova's case study utilized the use of multi agents in addressing the knowledge management model to achieve the alignment between human capital and business strategy in order to gain a sustainable competitive advantage in a highly competitive and dynamic market.

3.2 Second Case Study: Miettinen

Second case study that we have gone through is that of the Miettinen's case study. This case study has a centered focus on strengthening the relationship with their customers and customer orientation. There is a gap in the study where the service provider does not have a clear view on the exact customer needs and requirements. Hence, there is a high dependence on the leadership decision in designing the products and services to introduce it in the market. There is another gap that the customers cannot have clear service level agreements from the company hence indicating poor performance by the service provider towards its customer base. Although the company has undergone multiple efforts to improve the organizational

practices however, the workplace remains in a poor condition to reach to a positive position when it comes to customer satisfaction. The case study of Miettinen is evident to portray one of the service providers that have certainly failed to address the problem of managing the knowledge and hence reaching a maturity level of sustaining in a competitive market. The case study is evident in terms of how absence in deployment of knowledge management systems can cause a service provider to fail and loose in a highly competitive market.

4 Recommendations

We have presented in the previous section that various researchers have recommended the deployment of knowledge management systems in order to preserve organizational intellectual work. However, we believe that deploying solo elements of intellectual capital or innovative capacity or deployment of knowledge management system cannot address the overall organizational problem of labor mobility and professional employees' engagement and commitment. Hence, we recommend the use of a multi agent approach towards managing knowledge within a service provider. This will over time lead the service provider to be ahead of the competition and be a successful service provider one that is enduring, thriving and robust. The proposed knowledge management system is an active and dynamic platform that collects, classifies, disseminates, and supports the effective use of the continuous stream of knowledge that is produced within the service provider. To start with:

1. The service provider needs to identify all sources that produce knowledge within the organization. That is its professional employees, these can be the senior executive, middle managers and frontline workers at the organizational level. Where each layer has a certain role and responsibility towards producing organizational knowledge. The provider's stakeholders or known as external affiliated parties such as the regulators, vendors, suppliers, industry partners, customers, competitors, etc. Also, any other people element who the organization can benefit from their valuable insights and clues to develop organizational knowledge.
2. Next the service provider needs to identify the various competencies and expertise that are existent within the organization of its professional workers. This will help the provider to tap into its intellectual capital's tacit knowledge, values, expertise, experiences, ideas, insights, etc. and feeding this invaluable collected knowledge into the organizational enterprise resource planning system.
3. The service provider needs to also create the concept of *ba* as named by [34]. The concept of *ba* describes the "linkage points" of interactions and "where" they take place, as well as "when" and "how" (relationships). As such, *ba* can be interpreted as a shared context where people can interact with one another to

create new knowledge. In order to socialize, internalize, externalize and combine, organizations need to create platforms to enable these knowledge modes. Takeuchi suggested that these platforms can either be physical within or beyond the organizational boundaries by dedicating workspaces and meeting rooms or through virtual interacting platforms to enable various diverse members on knowledge community to interact and feed knowledge regardless of where (place) and when (time). Encouraging the creation of various taskforces and working groups internally within the organization or externally with service provider's affiliated external parties invokes healthy discussions in regards to listing ideas, insights and suggestions about the next emerging niche products and services or ways to create new markets to attract a new segment of customers.

4. Furthermore, to the multi agent approach towards managing knowledge as previous research has proven, to gain competitive advantage, organizations cannot rely merely on their employees and training systems to build competencies, knowledge and skills within the organization [35]. Organizations need to consider ways to fully exploit and capture existent tacit knowledge and expertise within the organization and convert it to organizational knowledge [1]. This is possible by effectively and efficiently capturing and converting the existent tacit knowledge and expertise of knowledge-based resources to organizational knowledge that is fed into the knowledge management system such as the provider's ERP system. There exist multiple ways to convert tacit knowledge to organizational knowledge through knowledge sharing, knowledge transfer, knowledge exchange, and knowledge creation. Numerous researches have shown the positive impact knowledge sharing, collaboration, and combination has towards reduction in production costs, faster completion of new product development projects, team performance, organizational innovation capabilities, and organizational performance [36].
5. Finally, the last piece to this proposed solution is integrating the SECI knowledge process model for knowledge sharing and transfer as named by [32]. This knowledge process emphasizes the two forms of knowledge that is tacit and explicit. Where any existing form of knowledge can be conveyed to the other form through knowledge sharing, transfer, and exchange with a 2 by 2 matrix. Hence, producing four distinct modes of knowledge sharing and this is named as the SECI knowledge process. Each of the four modes operates differently with a completely different outcome. The notion of SECI knowledge process is that of a continuous learning process. Hence, it is widely known as the spiral process. Organizations who value knowledge continuously initiate and sustain the SECI knowledge spiral process.

The implementation of such solution of using not one but multiple other alternatives and mechanisms will continuously ensure committed engagement and involvement of the service provider's professional employees. Hence, impeding the current problems of labor mobility and lack of employees' commitment.

5 Implementation Plan

After detailing out in the previous sections the two case studies and our recommendations. In this section we will explain the implementation plan to go about in order to set the way forward for a successful thought out phased approached implementation plan for a knowledge management system that best fits our target organization. The approach the target organization should use is as follows:

1. First, identifying the different levels of expertise, experience and competencies of the intellectual capital in the organization and that to be structured as follows: novice, advanced beginner, competent, proficient, and expert. These can be categorized and structured based on the number of years of practice, qualification, problems solved, interaction with others, contribution in the organization and seniority at workplace;
2. Second, identifying the knowledge agents within the knowledge process and these are the individual, the groups, the organizational and inter-organizational levels. These are the various levels of that contribute as feeds to the knowledge process;
3. Third, identifying the various stakeholders that contribute to the organizational knowledge and these are the external affiliated parties such as the vendors, suppliers, industry partners, regulators, customers and competitors and any other people element that can contribute to the organizational knowledge;
4. Forth, identifying the different factors of the organization that favors the knowledge and these are the leadership, culture, style, strategy, structure, management of intellectual capital, and information and communication systems;
5. Fifth, identifying the business focus areas of the organization starting with high priority to lowest priority and these can be: E-learning, Knowledge warehouses, corporate intranets, collaborative tools, intelligent agents, customer intelligence, process management, project tracking, collaborative platforms, etc.;
6. Sixth, ensuring to deploy the SECI spiral knowledge process within the overall framework of the knowledge management system where both tacit and explicit knowledge is transferred and converted into organizational intellectual knowledge;
7. Seventh, ensuring the concept of *ba* is deployed to ensure knowledge is equally shared, stored, exchanged, transferred, used, and spread across all different levels within the organization; and
8. Finally, ensuring that the overall integrated multi-agent knowledge management system has the ability to generate invaluable organizational intellectual knowledge that can create new markets, respond rapidly to customers, quickly develop new products and services, and dominate emerging markets.

6 Conclusion

To conclude, this study proposed the use of a multi agent approach as a knowledge management system in a technology service provider environment. The proposed solution has been developed after a thorough exploration of the use of state-of-the-art technologies such as the enterprise resource planning (ERP) system and knowledge management collaborative tools in order to enable seamless and rapid interaction amongst knowledge-based resources and dynamic organizational business processes enabling effective and efficient utilization of organization's cost, resources and development of quality and niche products and services. The target organization for this study was a technology service provider (TSP). That highly relies on expert intellectual professionals who can provide the required services to the clients. Hence, the retention of intellectual capital for a service provider is crucial to succeed and sustain in a competitive technology market. Labor mobility presents a problem for service providers who depends mainly on knowledge workers to provide its core business services. As when these knowledge-based resources leave their jobs they take the knowledge of organizational intellectual work with them. Hence, it is crucial for these service providers to deploy necessary solution, processes and mechanisms to safeguard, preserve and manage the organizational knowledge and knowhow to address the labor mobility problem. This study detailed out the implementation plan for the proposed solution of the use of a multi agent approach as a knowledge management system that addressed the labor mobility problem by integrating multiple dimensions and factors to allow all knowledge workers within the service provider to have the full potential of organizational intellectual knowledge. The access to organizational intellectual knowledge will give the knowledge workers the ability to collect, classify, disseminate, and support the effective and efficient use of the organizational intellectual knowledge to respond rapidly to customers, create new markets, quickly develop new products and services, and dominate emergent technologies.

Acknowledgements This work is a part of a project submitted to The British University in Dubai.

References

1. Nonaka I, Takeuchi H (1995) The knowledge-creating company: How Japanese companies create the dynamics of innovation. Oxford University Press
2. Nonaka I (1994) A dynamic theory knowledge of organizational creation. Organ Sci 5(1):14–37
3. Hunt SD (1995) The resource-advantage theory of competition: toward explaining productivity and economic growth. J Manag Inq 4(4):317–332
4. Grant RM (1996) Toward a knowledge-based theory of the firm. Strateg Manag J 17 (S2):109–122
5. Alameeri K, Alshurideh M, Al Kurdi B, Salloum SA (2021) The effect of work environment happiness on employee leadership, vol 1261. AISC 2021

6. Almazrouei FA, Alshurideh M, Al Kurdi B, Salloum SA (2020) Social media impact on business: a systematic review. In: International conference on advanced intelligent systems and informatics, pp 697–707
7. AlGhanem H, Shanaa M, Salloum S, Shaalan K (2020) The role of KM in enhancing AI algorithms and systems. Adv Sci Technol Eng Syst J 5(4):388–396
8. Spender J, Grant RM (1996) Knowledge and the firm: overview. Strateg Manag J 17(S2):5–9
9. Salloum SA, Al-Emran M, Shaalan K (2018) The impact of knowledge sharing on information systems: a review. In: 13th International conference, KMO 2018
10. Al Mansoori S, Salloum SA, Shaalan K (2021) The impact of artificial intelligence and information technologies on the efficiency of knowledge management at modern organizations: a systematic review. In: Recent advances in intelligent systems and smart applications. Studies in systems, decision and control, vol 295. Springer, Cham
11. Areed S, Salloum SA, Shaalan K (2021) The role of knowledge management processes for enhancing and supporting innovative organizations: a systematic review. In: Recent advances in intelligent systems and smart applications. Studies in systems, decision and control, vol 295. Springer, Cham
12. Almansoori A, AlShamsi M, Salloum SA, Shaalan K (2021) Critical review of knowledge management in healthcare, pp 99–119
13. Mehrez AAA, Alshurideh M, Kurdi BA, Salloum SA (2021) Internal factors affect knowledge management and firm performance: a systematic review, vol 1261. AISC 2021
14. Habeh O, Thekrallah F, Salloum SA, Shaalan K (2021) Knowledge sharing challenges and solutions within software development team: a systematic review. In: Al-Emran M, Shaalan K, Hassanien A (eds) Recent advances in intelligent systems and smart applications. Studies in systems, decision and control, vol 295. Springer, Cham
15. Alsharhan A, Salloum S, Shaalan K. The impact of elearning as a knowledge management tool in organizational performance
16. Davenport TH, Prusak L (1998) Working knowledge: How organizations manage what they know. Harvard Business Press
17. Liberona D, Ruiz M (2013) Analysis of knowledge management programs implemented in Chilean enterprises. Estud Gerenciales 29(127):151–160
18. Nonaka I, Von Krogh G (2009) Perspective—Tacit knowledge and knowledge conversion: controversy and advancement in organizational knowledge creation theory. Organ Sci 20 (3):635–652
19. Kivijarvi H (2008) Aligning knowledge and business strategies within an artificial ba context. In: Knowledge management and business strategies: theoretical frameworks and Empirical Research. IGI Global, pp 229–252
20. Bergeron B (2003) Essentials of knowledge management, vol 28. John Wiley & Sons
21. Hansen MT, Nohria N, Tierney T (1999) What's your strategy for managing knowledge. Knowl Manag Yearb 2000–2001 77(2):106–116
22. Takeuchi H, Nonaka I, Takeuchi H (2004) Hitotsubashi on knowledge management. John Wiley & Sons (Asia) Singapore
23. Alavi M, Leidner DE (2001) Knowledge management and knowledge management systems: conceptual foundations and research issues. MIS Q:107–136
24. Davenport TH, Völpel SC (2001) The rise of knowledge towards attention management. J Knowl Manag
25. Housel T, Bell AH (2001) Measuring and managing knowledge
26. Jiang J, Klein G (2000) Software development risks to project effectiveness. J Syst Softw 52 (1):3–10
27. Alavi M, Leidner DE (1999) Knowledge management and knowledge management systems: conceptual foundations and research issues. Insead Fontainebleau
28. Bernard J-G (2006) A typology of knowledge management system use by teams. In: Proceedings of the 39th annual Hawaii international conference on system sciences (HICSS'06), vol 7, pp 155a

29. Nonaka I, Toyama R (2005) The theory of the knowledge-creating firm: subjectivity, objectivity and synthesis. Ind Corp Chang 14(3):419–436
30. Uziene L (2010) Model of organization's intellectual capital measurement. Eng Econ 21(2)
31. Lohr S (2002) Gazing into 2003: economy intrudes on dreams of new services. New York Times, vol 3
32. Nonaka I, Takeuchi H (1995) The knowledge-creating company: how Japanese companies create the dynamics. Oxford University Press, USA
33. Cordova F, Gutierrez F (2018) Knowledge management system in service companies. Procedia Comput Sci 139(2018):392–400
34. Nonaka I, Takeuchi H (2007) The knowledge-creating company. Harv Bus Rev 85(7/8):162
35. Brown JS, Duguid P (2000) Balancing act: how to capture knowledge without killing it. Harv Bus Rev 78(3):73–80
36. Mesmer-Magnus JR, DeChurch LA (2009) Information sharing and team performance: a meta-analysis. J Appl Psychol 94(2):535

Impact of Knowledge Management on Organizational Performance

Reem Bayari, Arwa A. Al Shamsi, Said A. Salloum, and Khaled Shaalan

Abstract This paper aims to explore the relationship between knowledge management (KM) and organizational performance (OP). Systematic literature review was conducted to look at different practices for KM in different sectors and locations to build a good background about the relation between KM and OP. Many of researchers investigated the impacts and benefits of knowledge management. There is a limited literature review on a large scale in different countries and sectors to confirm the relation between the Knowledge Management and organizational performance. The result shows that ALL the studies found a significant positive relation between KM and OP. The best database resource for data collection is Emerald database. Quantitative method was the most method utilized in examining the relationship between KM and OP using the survey approach to collect data. The results displayed that most of the studies considered in the systematic review conducted in Pakistan. The participants in most of the conducted studies are employees.

Keywords Knowledge management · Organizational performance · Systematic literature review · Quantitative method

R. Bayari
Higher College of Technology, The British University in Dubai, Dubai, UAE

R. Bayari · A. A. Al Shamsi · K. Shaalan
Faculty of Engineering and IT, The British University in Dubai, Dubai, UAE

S. A. Salloum (✉)
School of Science, Engineering, and Environment, University of Salford, Salford, UK
e-mail: ssalloum@sharjah.ac.ae

S. A. Salloum
Machine Learning and NLP Research Group, Department of Computer Science, University of Sharjah, Sharjah, UAE

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

M. Al-Emran et al. (eds.), *Proceedings of International Conference on Emerging Technologies and Intelligent Systems*, Lecture Notes in Networks and Systems 322,
https://doi.org/10.1007/978-3-030-85990-9_82

1 Introduction

Over the past several decades there is an intensive works on the knowledge management such as analyzing the factors, dimensions, strategies, etc. [1]. The rapid changing in the performance's standards for the business environment shows that organization performance has become a key asset for any firms [2–4]. In [5] stated that Knowledge management is a tool that has the ability to facilitate the flow and sharing of information within an efficient way. However, the organization's ability to manage their assets and resources such as knowledge shows a lot of benefits for the business environments such as a knowledge sharing, organizing and improving the organization performance [6]. Therefore, investigation on the relation between knowledge management and organization performance is a critical issue. By controlling and managing knowledge assets, an enterprise can improves its level and increase its chances of success and competition with the success organizations. Knowledge management is an old topic for researchers. Reference [7], many enterprises stated improved organizations' performance depends not only on the assets, investments and strong strategies but also on the strength and effective knowledge management. Many of researchers investigated the impacts and benefits of knowledge management. There is a significant gap in the literature on large scale and different countries to confirm the relation between the Knowledge Management and organizational performance. Systematic reviews have been conducted in order to answer the following research questions:

RQ1: Does knowledge management impact on the organization's performance?
RQ2: How knowledge management affects organization's performance (Positively/negatively/neutral)?
RQ3: What are the main research methods that utilized in measuring the relationship between Knowledge management and organization's performance in the collected research papers?
RQ4: In which country the study is implemented and what are the kinds of the participants in the study?
RQ5: What are the best databases offered research papers in the context of the relation between knowledge management and performance?

2 Literature Review

The relationship between KM and organizational performance is considered by managers as a critical issue to the management scholars in different sectors such as industry and trading even in education. Reference [8], One of the fundamental objectives for KM is to improve the competitiveness of the firms by utilizing their existing knowledge, thus firms can stand in a competitive environment. Yuan et al. [9], Investigated the relationship between knowledge management and organizational

performance in enterprise by consideration two dimensions which are KM process and the impact environmental factors and found both dimensions have positive impact on the organization's performance. In addition, Ref. [10], Knowledge Management has a direct impact on business strategies formulation and implementation and KM lead to improve organizational performance by enhancing communication, collaboration, learning and adaption. Reference [11], Knowledge Management has significant and positive impact on IT Investment Success which highly impact organizational performance and the author achieved the result by using the quantitative methods utilizing the survey. In addition, Ref. [12], utilizing questionnaire approach shows Total Quality Management practices such as leadership, customer focus, teamwork, process management, and information and analysis had a weak positive impact on organizational performance, while Knowledge Management processes have positive impact on organizational performance. Shahzad et al. [13], applied his study in Pakistan and targeted managers in the business organizations sector. The study indicates that KM strategy has a significant positive impact on organizational creativity and performance. Mageswari et al. [14], targeted the manufacturing companies and found Knowledge Management presented significant positive impact on performance measures. Jyoti and Rani [15], found High performance work system and organizational performance impacted positively by Knowledge management. Reference [16], knowledge management is essential to improve organizations performance. Reference [17], Knowledge Management processes i.e. knowledge creation, knowledge capture and storage, knowledge sharing, and knowledge application and use impacted positively and significantly "operational, quality and innovation performance of public sector in the UAE. Reference [18], KM has a significant positive effects in growing and learning dimension otherwise it is insignificant [19], found the KM have positive effects on the human resource performance. Reference [20], in Iran, KM have a positive effects over the organization's productivity. References [21, 22] KM in term of process has stronger link with the OP comparable to organization infrastructure. Reference [23], Knowledge management lead to get a unique performance. References [24, 25], show a strong positive relationship between KM and innovation's performance. References [26, 27], examined the impact of KM infrastructure and practices on the banking sectors and found a strong positive relationship between. According to the results from this literature that focused on exploring the relationship between KM and OP in different sectors and countries. There is a gap of this measuring in some sectors and countries. Most researchers work targeted the employee and managers. The majority of studies applied on the banking sector while some sectors only has one or two studies.

3 Methods

An extensive literature review is a critical phase before doing any research study [28]. It helps in establishing a good background for knowledge accumulation, in order to enables the theories' extensions and developments, find and fill the gaps,

… etc. [29]. A systematic literature review is a kind of literature review but more organized, based on a certain criteria and based on explicit research questions, [30]. In this systematic review, the review was conducted in four distinct phases: the identification of inclusion and exclusion criteria, data sources and search strategies, quality assessment, and data coding and analysis. The details of these phases are demonstrated in the following subsections.

3.1 Inclusion and Exclusion Criteria

3.1.1 Inclusion Criteria

- Should involve knowledge management and knowledge performance
- Should involve an impact or influence or role or effects.
- Should be written in English.
- Should have methods such as experiments or survey or interview or (interview and survey)
- The publication dates should be within the period from 2014 to 2019.

3.1.2 Exclusion Criteria

- Knowledge management strategies, knowledge sharing, security/processes/factors/processes or factors
- Any study that not reflecting the relation between knowledge management and level of performance.

3.2 Data Sources and Search Strategies

This study based on systematic review, utilize six sub-sequent databases to collect the research articles. The involved in the articles were collected after an extensive search of latest studies via different databases: Emerald insight, IEEE, ScienceDirect, Springer, Wiley, and Google Scholar. The search of these studies was in October, November 2019. The search terms include the keywords "knowledge management" and "organization/firms" performance". The selection of the keywords is an essential step in any systematic review as it determines which articles are to be retrieved [31]. Using the aforementioned keywords, the total of retrieved research Articles is 101 articles. 7 articles were excluded because it is duplicates articles, 61 articles were removed after screening the abstract and conclusion, the authors confirmed after assessing full text total number of remaining

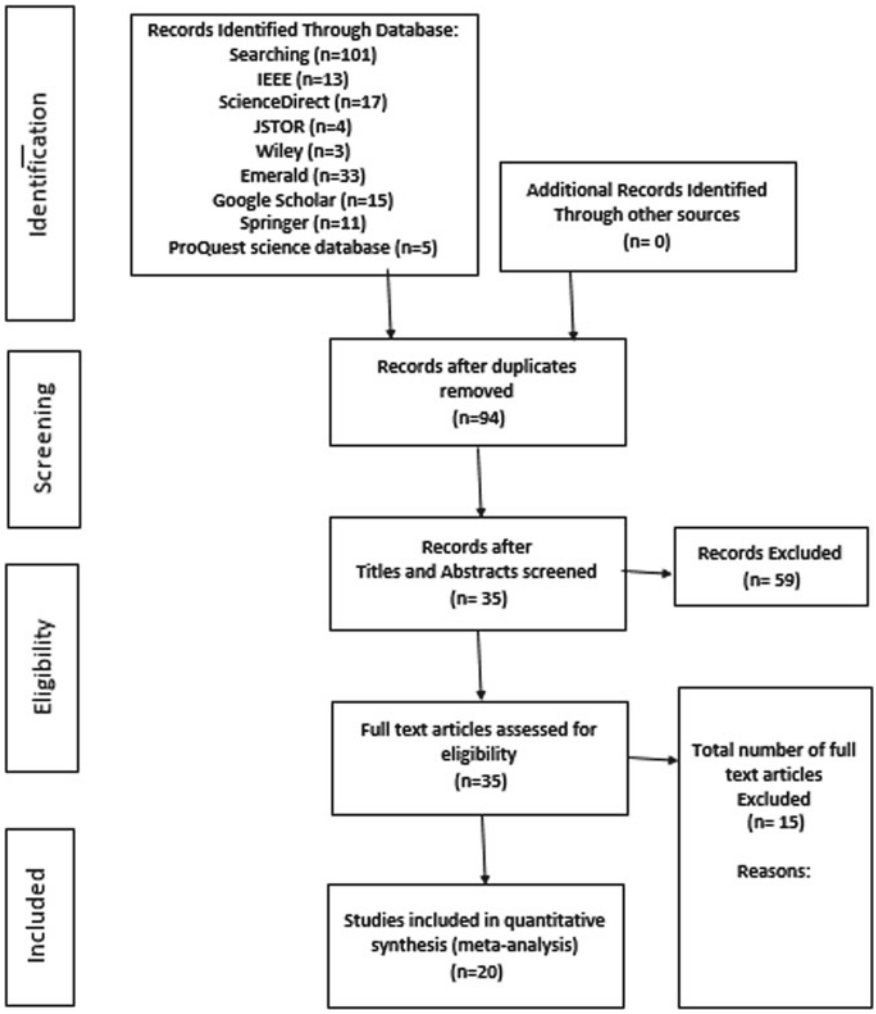


Fig. 1 PRISMA flow diagram

articles is 20. The finding and filtering research work in this study is followed developed as (PRISMA)) in [32] (Fig. 1).

4 Results

Reviewing the 20 research papers considered in this systematic review that have been published since 2014, the findings of this systematic review have been stated based up on the six research questions.

4.1 RQ1: Does Knowledge Management Impact on the Organization's Performance?

Number of research papers and studies have been carried out to investigate whether the Knowledge Management have an impact on the Organization's Performance. In this systematic review authors reviewed 101 research papers and studies have been screened and only 20 research papers and studies considered in this systematic review. These 20 research papers or studies explored to find whether the Knowledge Management have impact on the Organization's performance. Figure 2 indicated that Organization's Performance affected by Knowledge Management.

Figure 2 presents that all the research papers and studies considered in this systematic review agree that the Knowledge Management have impact on the Organization's Performance.

4.2 RQ2: How Knowledge Management Affects Organization's Performance (Positively/Negatively/Neutral)?

Systematic review methods have been applied on number of research papers and studies that considered the impact of knowledge management on organizational performance. The main concern of the authors of this study is to investigate how the knowledge management would impact the organizational performance. After analyzing selected research papers and studies, it is clear that all of the selected research papers as well as studies concluded that the knowledge management positively impacted the organizational performance. This result is represented in Fig. 3.

Fig. 2 Knowledge management impact on the organization's performance

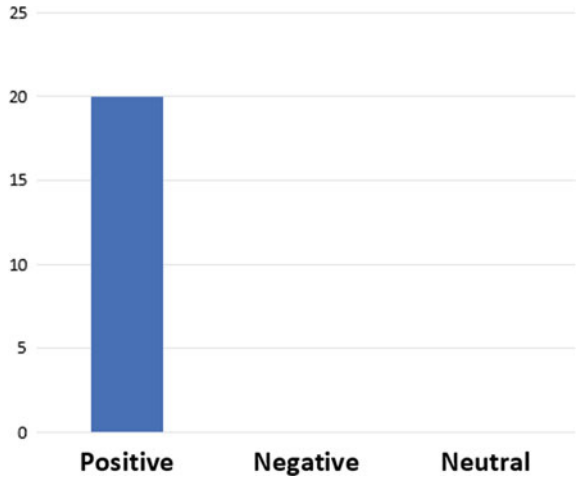


Fig. 3 Knowledge management impacts on organization's performance

This result is aligned along with results from literature review that stated that knowledge management has significant and positive impact on organizational performance.

4.3 RQ3: What Are the Main Research Methods that Utilized in Measuring the Relationship Between Knowledge Management and Organization's Performance in the Collected Research Papers?

When reviewing the selected research papers for the systematic review, the methods utilized for measuring the relationship between knowledge management and organization's performance have been reported. The Fig. 4 below represents that 95% of the analyzed research papers rely on questionnaire surveys (N = 19) for data collection, while only 5% of the selected research papers rely on both surveys and questionnaire for data collection.

4.4 RQ4: In Which Country the Study is Implemented and What Are the Kinds of the Participants in the Study?

The studies and research papers considered in this systematic review are analyzed and the following results are reported as presented in Fig. 5. The results displayed that most of the studies considered in the systematic review conducted in Pakistan

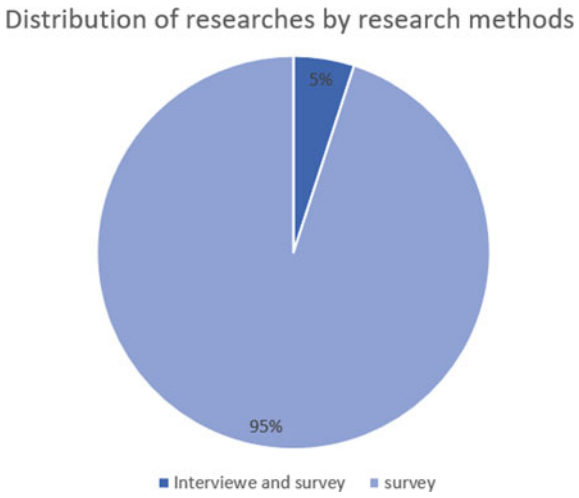


Fig. 4 Distribution of researches by research methods used for data collection

with 4 studies, followed by Iran with 3 studies. After that India, Jordan and Palestine with 2 studies for each of them, while for the remaining countries only one study or research paper was considered in this systematic review. It is important to mention that all of the studies considered in this research paper indicated positive impact of knowledge Management on the Organizational Performance as mentioned earlier.

Statistics of the participants in the studies and research papers considered in this systematic review are presented in Fig. 6. It is obvious that most of the conducted studies considered all the employees (N = 10), 6 studies targeted managerial

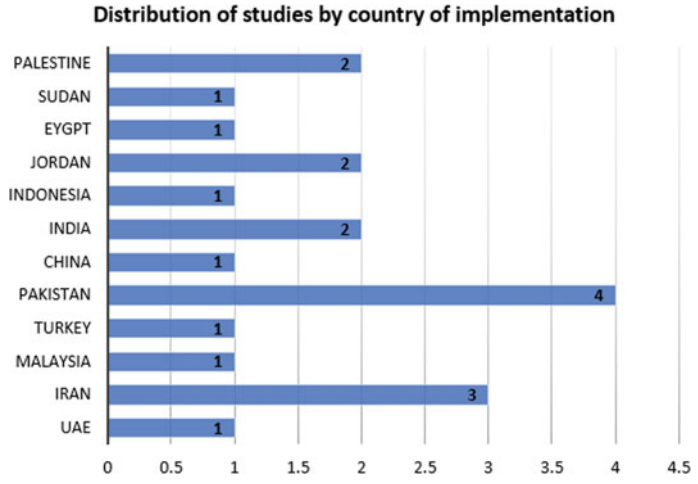


Fig. 5 Distribution of studies by country of implementation

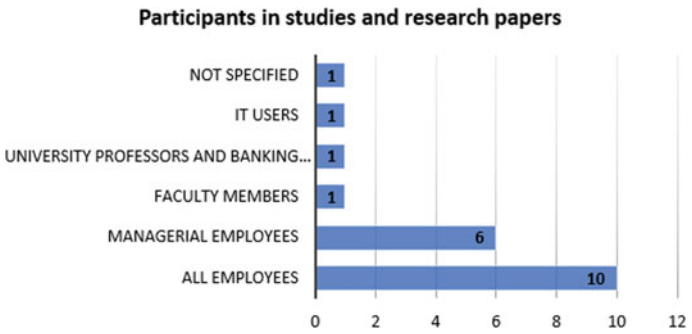


Fig. 6 Participants in studies and research papers

employees, while IT users, faculty members and University professors and banking experts participated only in one study or research paper from the studies considered in this systematic review.

4.5 RQ5: What Are the Best Databases Offered Research Papers in the Context of the Relation Between Knowledge Managements and Performance?

Statistics after reviewing studies and research papers considered in this systematic review represented the best databases that offered research papers in the context of the relation between knowledge managements and performance. Figure 7 represented that the best database is Emerald database (N = 7) followed by Google Scholar (N = 6), after that IEEE database and ScienceDirect both of them offered 3 studies or research papers (N = 3) and lastly Springer database with only one research paper.

5 Conclusion

The study focused on exploring the researchers work about the nature of the relationship between KM and OP. Authors systematically review researches and articles that investigated the relationship between KM and OP. Authors found that applying KM positively impacted performance of the organizations in all the research papers in a different sector. Most researchers utilized the quantitative methods in examining the relationship between KM and OP. There is a slight difference between the best databases to collect data about the impact of KM on OP which are Emerald Insight and Google Scholar. Authors reported that a limited literature review on a large scale in different countries and sectors to confirm the

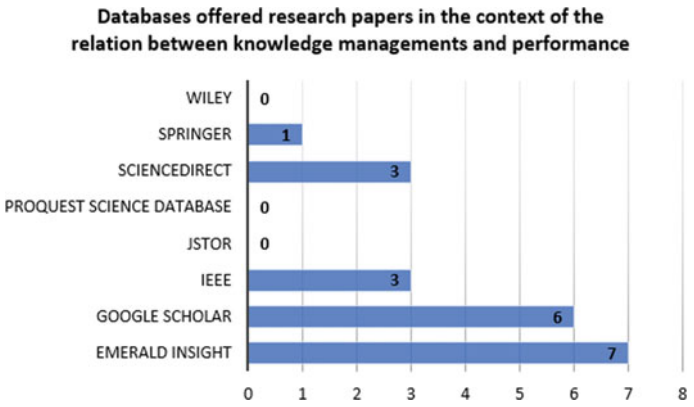


Fig. 7 Databases offered research papers in the context of the relation between knowledge managements and performance

relation between the Knowledge Management and organizational performance. For future work, authors recommended an investigation on the KM standards, the factors that affect and limit the efficiency of KM.

Acknowledgements This work is a part of a project submitted to The British University in Dubai.

References

1. Ooi K-B (2014) TQM: a facilitator to enhance knowledge management? A structural analysis. Expert Syst Appl 41(11):5167–5179
2. Ooi K-B, Lin B, Teh P-L, Chong AY-L (2012) Does TQM support innovation performance in Malaysia's manufacturing industry? J Bus Econ Manag 13(2):366–393
3. Ooi K, Arumugam V, Teh P, Chong AY (2008) TQM practices and its association with production workers. Ind Manag Data Syst
4. Ooi K, Teh P, Chong AY (2009) Developing an integrated model of TQM and HRM on KM activities. Manag Res News
5. McAdam R, McCreedy S (1999) A critical review of knowledge management models. Learn Organ
6. Choi B, Poon SK, Davis JG (2008) Effects of knowledge management strategy on organizational performance: a complementarity theory-based approach. Omega 36(2): 235–251
7. Fattahiyan S, Hoveida R, Siadat SA, Talebi H (2013) The relationship between knowledge management enablers, processes resources and organizational performance in Universities (Case study: selected Universities of the Isfahan Province). Int J Educ Res 1(11):1–14
8. Chang CL, Lin T-C (2015) The role of organizational culture in the knowledge management process. J Knowl Manag
9. Yuan Y, Chai H, Liu L (2017) How knowledge management impacts performance: An empirical study in Chinese knowledge-intensive enterprises. In: 2017 IEEE international conference on industrial engineering and engineering management (IEEM), pp 1103–1107

10. Eltayeb S, Kadoda G (2017) Business strategies and organizational performance. In: Sudan conference on computer science and information technology (SCCSIT)
11. Putra DM, Winoto DW (2017) The role of knowledge management in the success rate of it investment and its impacts on the organization performance a survey in the ministry agencies, local governments, universities and banks in Indonesia. In: 2017 3rd International conference on science in information technology (ICSITech), pp 150–155
12. Almahamid SM, Qasrawi ST (2017) The impact of TQM practices and KM processes on organisational performance. Int J Qual Reliab Manag
13. Shahzad K, Bajwa SU, Siddiqi AFI, Ahmid F, Sultani AR (2016) Integrating knowledge management (KM) strategies and processes to enhance organizational creativity and performance: an empirical investigation. J Model Manag
14. Mageswari SDU, Sivasubramanian RC, Dath TNS (2017) A comprehensive analysis of knowledge management in Indian manufacturing companies. J Manuf Technol Manag
15. Jyoti J, Rani A (2017) High performance work system and organisational performance: role of knowledge management. Pers Rev
16. Shehata GM (2015) Leveraging organizational performance via knowledge management systems platforms in emerging economies. VINE
17. Singh SK, Al Ahbabi SA (2018) Employee perception of impact of knowledge management processes on public sector performance
18. Valmohammadi C, Ahmadi M (2015) The impact of knowledge management practices on organizational performance: a balanced scorecard approach. J Enterp Inf Manag 28(1): 131–159
19. Torabi MHR, Kyani A, Falakinia H (2016) An investigation of the impact of knowledge management on human resource performance in management of Keshavarzi bank branches in Tehran. Procedia-Social Behav Sci 230:471–481
20. Torabi F, El-Den J (2017) The impact of knowledge management on organizational productivity: a case study on Koosar Bank of Iran. Procedia Comput Sci 124:300–310
21. Alaarj S, Abidin-Mohamed Z, Bustamam USBA (2016) Mediating role of trust on the effects of knowledge management capabilities on organizational performance. Procedia-Social Behav Sci 235:729–738
22. Al-Hayaly MAM, Alnajjar FJS (2016) Knowledge management processes and their impact on organizational performance, the adoption balanced scorecard: the moderating role of quality assurance standards—an applied study on private Jordanian Universities. Int J Bus Manag 11(6):70
23. Abu-Naser SS, Al Shobaki MJ, Abu Amuna YM (2016) Knowledge management maturity in universities and its impact on performance excellence "Comparative study"
24. Nawaz MS, Shaukat S (2014) Impact of knowledge management practices on firm performance: testing the mediation role of innovation in the manufacturing sector of Pakistan. Pakistan J Commer Soc Sci 8(1):99–111
25. Kamasak R, Yavuz M, Altuntas G (2016) Is the relationship between innovation performance and knowledge management contingent on environmental dynamism and learning capability? Evidence from a turbulent market. Bus Res 9(2):229–253
26. Ahmed S, Fiaz M, Shoaib M (2015) Impact of knowledge management practices on organizational performance: an empirical study of banking sector in Pakistan. FWU J Soc Sci 9(2):147–167
27. Imran MK (2014) Impact of knowledge management infrastructure on organizational performance with moderating role of KM performance: an empirical study on banking sector of Pakistan. Inf Knowl Manage 4(8):85–98
28. Al-Emran M, Mezhuyev V, Kamaludin A, Shaalan K (2018) The impact of knowledge management processes on information systems: a systematic review. Int J Inf Manage 43:173–187
29. Marangunić N, Granić A (2015) Technology acceptance model: a literature review from 1986 to 2013. Univers Access Inf Soc 14(1):81–95

30. Khan KS, Kunz R, Kleijnen J, Antes G (2003) Five steps to conducting a systematic review. J R Soc Med 96(3):118–121
31. Costa V, Monteiro S (2016) Key knowledge management processes for innovation: a systematic literature review. VINE J Inf Knowl Manage Syst
32. Moher D et al (2015) Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 4(1):1