Sridaran Rajagopal · Kalpesh Popat ·
Divyakant Meva · Sunil Bajeja ·
Pankaj Mudholkar (Eds.)

Communications in Computer and Information Science 2429

Artificial Intelligence Based Smart and Secured Applications

Third International Conference, ASCIS 2024
Rajkot, India, October 16–18, 2024
Revised Selected Papers, Part VI

Part 6

Springer
ASCIS 2024
Marwadi University
Marwadi Chandarana Group
NAAC A+

Communications in Computer and Information Science 2429

Series Editors

Gang Li, *School of Information Technology, Deakin University, Burwood, VIC, Australia*
Joaquim Filipe, *Polytechnic Institute of Setúbal, Setúbal, Portugal*
Zhiwei Xu, *Chinese Academy of Sciences, Beijing, China*

Rationale
The CCIS series is devoted to the publication of proceedings of computer science conferences. Its aim is to efficiently disseminate original research results in informatics in printed and electronic form. While the focus is on publication of peer-reviewed full papers presenting mature work, inclusion of reviewed short papers reporting on work in progress is welcome, too. Besides globally relevant meetings with internationally representative program committees guaranteeing a strict peer-reviewing and paper selection process, conferences run by societies or of high regional or national relevance are also considered for publication.

Topics
The topical scope of CCIS spans the entire spectrum of informatics ranging from foundational topics in the theory of computing to information and communications science and technology and a broad variety of interdisciplinary application fields.

Information for Volume Editors and Authors
Publication in CCIS is free of charge. No royalties are paid, however, we offer registered conference participants temporary free access to the online version of the conference proceedings on SpringerLink (http://link.springer.com) by means of an http referrer from the conference website and/or a number of complimentary printed copies, as specified in the official acceptance email of the event.

CCIS proceedings can be published in time for distribution at conferences or as post-proceedings, and delivered in the form of printed books and/or electronically as USBs and/or e-content licenses for accessing proceedings at SpringerLink. Furthermore, CCIS proceedings are included in the CCIS electronic book series hosted in the SpringerLink digital library at http://link.springer.com/bookseries/7899. Conferences publishing in CCIS are allowed to use Online Conference Service (OCS) for managing the whole proceedings lifecycle (from submission and reviewing to preparing for publication) free of charge.

Publication process
The language of publication is exclusively English. Authors publishing in CCIS have to sign the Springer CCIS copyright transfer form, however, they are free to use their material published in CCIS for substantially changed, more elaborate subsequent publications elsewhere. For the preparation of the camera-ready papers/files, authors have to strictly adhere to the Springer CCIS Authors' Instructions and are strongly encouraged to use the CCIS LaTeX style files or templates.

Abstracting/Indexing
CCIS is abstracted/indexed in DBLP, Google Scholar, EI-Compendex, Mathematical Reviews, SCImago, Scopus. CCIS volumes are also submitted for the inclusion in ISI Proceedings.

How to start
To start the evaluation of your proposal for inclusion in the CCIS series, please send an e-mail to ccis@springer.com.

Sridaran Rajagopal · Kalpesh Popat ·
Divyakant Meva · Sunil Bajeja ·
Pankaj Mudholkar
Editors

Artificial Intelligence Based Smart and Secured Applications

Third International Conference, ASCIS 2024
Rajkot, India, October 16–18, 2024
Revised Selected Papers, Part VI

Springer

Editors
Sridaran Rajagopal
Marwadi University
Rajkot, Gujarat, India

Kalpesh Popat
Marwadi University
Rajkot, Gujarat, India

Divyakant Meva
Marwadi University
Rajkot, Gujarat, India

Sunil Bajeja
Marwadi University
Rajkot, Gujarat, India

Pankaj Mudholkar
Marwadi University
Rajkot, Gujarat, India

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-3-031-86304-2 ISBN 978-3-031-86305-9 (eBook)
https://doi.org/10.1007/978-3-031-86305-9

© The Editor(s) (if applicable) and The Author(s), under exclusive license
to Springer Nature Switzerland AG 2025

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

If disposing of this product, please recycle the paper.

Preface

"A Little Progress Each Day, Adds up to a Big Result"

Our 3rd International Conference on Advancements in Smart Computing & Information Security (ASCIS 2024) invited original research papers under five tracks, namely, **AI & ML, Smart Computing, Cyber Security, Networking and Cloud Computing and Computer Applications for Sustainability.** We received 667 submissions and 181 were recommended for publication in Springer's CCIS series. The papers will appear as 6 volumes (Volume Numbers ranging 2424–2429). We are very much thankful to the editorial board of **Springer CCIS** for their continuous support.

The submitted papers were involved in an open review process, involving three reviewers for each paper. We are thankful to the Technical Program Committee (TPC) members who were very supportive in the review process.

Several sponsors who contributed generously towards the successful running of ASCIS 2024, including **Science and Engineering Research Board (SERB), K7 International and D- Link Academy**. These collaborations enriched the conference and extended its outreach.

ASCIS 2024 was conducted during 16–18 October, 2024, and featured insightful keynote addresses from national and international experts in their respective fields. The line-up included **Prabir Kumar Biswas** from IIT Kharagpur, Cybersecurity expert **Ram Kumar G, Arnav Bhavsar** from IIT Mandi, **Murali Raman** from Asia Pacific University, and **Aravind Benegal** from NLB Services.

The keynotes addressed several recent technological advancements whereas the hands-on workshops provided practical exposure to the students and participants. The PhD forum featuring the research works of research scholars and recent PhD graduates added value to the conference and acted as a platform for knowledge sharing.

These volumes titled **"Artificial Intelligence based Smart & Secured Applications"** are the consolidation of all such solutions employing technologies of AI, ML, Smart Computing and, Cybersecurity in every significant field seeking research solutions.

October 2024 Sridaran Rajagopal

Organization

General Chair

R. Sridaran	Marwadi University, India

Program Committee Chairs

R. Sridaran	Marwadi University, India
Kalpesh Popat	Marwadi University, India
Divyakant Meva	Marwadi University, India
Sunil Bajeja	Marwadi University, India
Pankaj Mudholkar	Marwadi University, India

Steering Committee

Ketanbhai Marwadi	Marwadi University, India
Jitubhai Chandarana	Marwadi University, India
P. Bhanu Prasad	Future Technology
R. B. Jadeja	Marwadi University, India
Sanjeet Singh	Marwadi University, India
Prabir Kumar Biswas	IIT Kharagpur, India
Anil Kumar Sao	IIT Bhilai, India
Arnav Bhavsar	IIT Mandi, India
Naresh Jadeja	Marwadi University, India
Shamala Subramanyam	Universiti Putra Malaysia, Malaysia
Padmavathi	Avinashilingam Institute for Home Science and Higher Education for Women, India
T. Devi	Bharathiar University, India
K. Iyakutti	SRM Institute of Science and Technology, India
Ram Kumar G.	Global Automotive Company, India

Technical Program Committee Chairs

AI and ML Track

Hitesh Chhinkaniwala	Adani University, India
Ahlad Kumar	National Forensic Sciences University, India
Mahendra Kumar Gourisaria	Kalinga Institute of Industrial Technology, India

Cyber Security Track

Ruchira Naskar	Indian Institute of Technology, Kharagpur, India
Shravani Shahapure	Deloitte Touche Tohmatsu, India
Ritika Vivek Ladha	Adani University, India

Smart Computing Track

Manish Khare	Dhirubhai Ambani Institute of Information and Communication Technology, India
Dinesh Kumar	Bennett University, India
Rebakah Geddam	Nirma University, India

Networking and Cloud Computing Track

Rachna Jain	Bhagwan Parshuram Institute of Technology, India
Santosh Kumar Majhi	Guru Ghasidas Vishwavidyalaya (Central University), India
Nishant Doshi	Pandit Deendayal Energy University, Gandhinagar, India

Computer Applications for Sustainability Track

Manas Ranjan Pradhan	Skyline University College, UAE
J. Ramkum	Krishna Arts and Science College, India

Additional Reviewers

Arun Raj Lakshminarayanan
Dhanamma S. Jagli
Tushar H. Jaware
Kamal Sutaria
Mahalakshmi G.
Nebojsa Bacanin
Rahul Ashok Patil
Satvik Khara
Vijay D. Katkar
Vipin Sharma
Amarinder Kaur
Anita Gautam Khandizod
Bhoopendra Singh
Delecta Jenifer Rajendren
Harishchander Anandaram
Karuna Nidhi Pandagre
Amrita Kumari
Dipti H. Domadiya
Prashant Sen
Rajasekaran Selvaraju
Saran Raj Sowrirajan
Harshal Anil Salunkhe
Sumathy S.
Hemant Ingale
Jamberi K.
Jayashree Nair
Jyoti Kharade
Keerti Jain
Kumaresh N.
Mythili Shanmugam
Nikita Bhatt
Parvathaneni Naga Srinivasu
Praveen P. V. Sundar
Ravirajsinh S. Vaghela
Richa Adlakha
Ritesh Patel
Sharanyaa S.
Srinivasulu Raju S.
Swetta Kukreja
Vinothina V.
Ajay Kushwaha
Anbumani K.
Ashwin B. Makwana
Chandresh K. Kumbharana
Divyakant T. Meva
Raji C. G.
Anita M. Indu
Kailash Patidar
Kannadhasan Suriyan
Hardik K. Molia
Jayashri Pattar
Jonnadula Narasimharao
Kaviyarasi R.
Ramakrishna Garine (Ram)
Nageswari D.
Prema C.
Rajender Kumar
Ram Ratan
S. Amuth Sachin H. Gajjar
Sreenaga Chidambara Vinjamuri Dattu
Vipul Shah
A. Yovan Felix
Abhay Singh Bhadauria
Aditi Sharma
Ajay Kumar Kushwaha
Ajita Deshmukh
Akash Saxena
Amita Sharma
Ankit Subhash Didwania
Anubhav Kumar Prasad
Arun Adiththan
Asha V.
Ashokkumar Baldevbhai Prajapati
Avinash Sharma
Bhagvati Parekh
Brijesh R. Jajal
Dafni J. Rose
Dhivya K.
Dipti Chauhan
Divya R.
Divyanshu Chandra
Ashwin R. Dobariya
Prabhat Sharma
Rohit Kanauzia
Abhinav Tomar
Akshara Dave

Ashish Saini
Avnip Deora
Dimple M. Thakar
Gaurav Kumar Ameta
Jayant R. Nandwalkar
Kapil Joshi
Lokesh P. Gagnani
Nisha Khurana
Samir B. Patel
Sri Hari Nallamala
Sunil L. Bajeja
T. Sathish Kumar T.
Parwinder Kaur
Rajesh Bansode
Saraswathi S.
T. Buvaneswari Muruganantham
Vijayalakshmi P. S.
Galiveeti Poornima
Hari Kumar Palani
Hemraj S. Lamkuche
Jafar Ali Ibrahim Syed Masood
Jay A. Dave
Jebakumar Immanuel D.
Juliet Rozario
Kajal S. Patel
Karthikeyan Eswaramurthy
Kavitha Ganesh
Krishnakumar A.
Lata Suresh
Latchoumy P.
M. R. Ramesh
Madhu Shukla
Manisha Rawat
Maruthamuthu R.
Maulika Patel
Megha P. Mudholkar
Mohamed Iqbal M.
Mohit Tiwari
Muralidharan R.
Narayan Joshi
Neeru Sharma
Neha Sharma
Nidhi M. Patel
Nilu Singh
Pankaj K. Mudholkar
Parth Gautam
Pranav Verma
Praveen Kumar
Priyanka Sharma
Purnendu Bikash Acharjee
Radha B.
Rajesh Yadav
Rakesh Kumar Yadav
Renjith V. Ravi
Richa Ashok Modiyani
Rohit Goyal
Rupali Atul Mahajan
S. S. Priscila
Saifullah Khalid
Sandeep Mathur
Saswati Mahapatra
Senthil Kumar R.
Shamala K. Subramaniam
Shanti Verma
Shweta Sharma
Srinivasa Rao Pallapu
Subramanian Balambigai
Sudipta Hazra
Sunil Gautam
Suresh B. Kumar
Tanmay Kasbe
Thirumurugan Shanmugam
Uma N. Dulhare
Umesh S. Pinjarkar
Vaibhav C. Gandhi
Vandani Verma
Vipin Sharma
Vishvjit K. Thakar
Yogesh Ghodasara
Chandran C. P.
Raghu N.
Jaimin N. Undavia
Kavipriya P. P.
Mohamed Mosbah
Padma Selvaraj
Remegius Praveen Sahayaraj
Shadab Siddiqui
Vineet Kumar Singh
Abhishek Sharma
Anilkumar Suthar

Bharanidharan G.
Chandra J.
Disha J. Shah
John T. Abraham
P. Rizwan Ahmed
B. Surendiran
Dushyantsinh B. Rathod
Priya K.
Rushikumar R. Raval
Ajitha V.
Kajal Patel
Gourav Gupta
J. Grace Hannah
Jasminder Kaur Sandhu
Jinal H. Tailor
Karthik B.
Kruti K. Sutaria
K. P. Malarkodi
Naresh Kumar
Nirav Bhatt
Pradip Mathuradas Jawandhiya
Rajeswari S.
Revathy G.
Ripal D. Ranpara
Senthilkumar Meyyappan
Shikha Maheshwari
Sweeti Sah
Vijay Karnatak
A. Maheswary
Anamika Rana
Anjali Diwan
Biswaranjan Mishra
Chintan Patel
Himanshu K. Maniar
T. S. Murugesh
Darshankumar C. Dalwadi
Anitha K.
Priya Chandran
Jalpesh Vasa
Jeevitha R. R.
Jose M. Molina
Lilly Florence M.
Mohan Subramani
Naresh Kshetri
Qixia Zhang
Rajeswari M.
Roshan Anant Gangurde
Safvan Vahora
Vijaya Kittu Manda
Yugendra Devidas Chincholkar
A. P. Nirmala
Abhilasha Vyas
Ahmed Banimustafa
Ajay M. Patel
Ajitha I.
Akoramurthy Balasubramaniam
Angeline R.
Ankur N. Shah
Anup Palsokar
Aruna Animish Pavate
Ashish Kumar
Ashwin Raiyani
Balraj Verma
Boopathi Raja G.
Chandra Mohan
Darshita S. Pathak
Dipak Ramoliya
Divya Didwania
Anant G. Kulkarni
Chintan B. Thacker
Deepak Kumar Verma
Rashmi Soni
Saraswathi S.
Abhishek Sharma Padmanabhan
Ankit J. Faldu
Asmita Manna
Daxa Vekariya
Disha H. Parekh
Jay Kumar Jain
Jaypalsinh A. Gohil
Krupa S. Mehta
Nafees Akhter Farooqui
Premkumar Borugadda
Rupesh Kumar Jindal
Shadab Siddiqui
Sudhanshu Maurya
Sunil Gupta
Vatsal H. Shah
Rajan Patel
Ramesh T. Prajapati

Sreejith Vignesh B. P.
Veena Soni
G. Charles Babu
Gaurav Agrawal
Harish Padmanaban
Himanshu Rai
Jane K. Nithya
Jaydeep R. Ramani
Juhi Singh
Jyotirmoy Pathak
Kamal Saluja
Karthikeyan R.
Kedir Lemma Arega
Kumuthini C.
Lataben J. Gadhavi
Lipsa Das
M. Vinoth Kumar
Mahalakshmi G.
Manoj Kr. Mishra
Mastan Vali Shaik
Meet Patel
Minal S. Shukla
Mohammed Wajid Khan
Monther Ali Tarawneh
Nagaraju Kilari
Neha Parashar
Neha Ripal Soni
Nilesh Vijay Sabnis
Pagalla Bhavani Shankar
Paresh V. Virparia
Pragadeswaran S.
Prashant P. Pittalia
Preethi S. R.
Priyanka Suyal
Pushparaj Pal
Rajib Biswas
Ramkumar Jaganathan
Rinkoo Bhatia
Ruchika Pharswan
Rutvi R. Shah
Sadhana Singh
Samriti Mahajan
Sanjeevkumar D. Angadi
Savitha Jaganathan
Shafi Pathan
Shantha Mary Joshitta
Shilpa Mehta
Sivanesan Rajangam
Subhadip Sarkar
Sudha R.
Suhasini Vijaykumar
Sunil Saxena
Surya Kameswari Uduga
T. D. Sudhakar
Tarannum Bloch
Tripti Tiwari
Umang Thakkar
Vaibhav A. Gandhi
Vallidevi Krishnamurthy
Vikas Tripathi
Vipul A. Shah
Yogendra Kumar
Chetan Dudhagara

Our Sponsors

Sr. No.	Name of the Sponsor	Logo
1	SERB India	
2	D-Link	
3	Digital India	
4	K7 International	
5	Samatrix.io	
6	Marwadi Technologies	
7	Stelcore	

Abstracts of Keynotes

Transforming Industry 4.0 with Generative AI and Cloud Computing

Gundarapu Pavan

Safran DIFIT, France

Industry 4.0, characterized by the convergence of automation, IoT, and advanced analytics, is reshaping industries globally. As organizations strive for greater efficiency, innovation, and adaptability, the synergy between cloud computing and generative AI is emerging as a transformative force. This keynote will explore the vital role of cloud computing as the backbone of Industry 4.0, providing scalable infrastructure and real-time data accessibility. It will delve into what generative AI is, how it acts as a catalyst for innovation, and how it accelerates the development of intelligent systems, from predictive maintenance to automated design and decision-making. By examining the synergies between these two technologies, the keynote will illustrate their potential to revolutionize industrial operations and unlock new business value.

Attendees will gain insights into the tangible benefits and ROI of integrating cloud and AI solutions. The keynote will also address the challenges organizations face—ranging from data security to infrastructure optimization—and offer strategic considerations for successful implementation. Looking ahead, we'll explore future trends in Industry 4.0, where cloud-enabled generative AI could redefine the industrial landscape, driving sustainable growth and continuous innovation.

The session will conclude with a call to action, encouraging students and industry leaders to embrace these technologies to remain competitive in a rapidly evolving digital economy.

Brain Decoding with Deep Learning

Arnav Bhavsar

IIT, Mandi, India

The talk will focus on visual brain decoding via EEG using deep learning techniques. Brain decoding, also known as Perceptual Brain Decoding (PBD), involves using brain responses to different stimuli to identify the original stimulus or its characteristics.

Specifically, we target an external visual stimulus, which involves object images on a computer screen. The goal is to reconstruct the class of the image from the EEG data captured with observing the image. The decoding involves either classifying reconstructing the type of image shown during the visual task.

For reconstructing class-specific images, we will discuss the utilization of generative networks such as GANs, using embeddings from a pre-trained EEG classifier network as input. Additionally, we will also cover an end-to-end generative network (NeuroGAN) for image synthesis, where the generator produces class-specific embeddings and images from EEG signals.

Advanced Insights into Deep Fake Recognition

Deepak Kumar Jain

Dalian University of Technology, China

The rapid advancement of artificial intelligence and machine learning has led to the proliferation of deep fake technology, enabling the creation of highly realistic, AI-generated synthetic media. While deep fakes offer exciting possibilities in fields like entertainment and virtual reality, they also pose significant threats to public trust, cyber security, and political stability. This presentation, *Advanced Insights into Deep Fake Recognition*, explores the complex techniques used to produce deep fakes and highlights the growing challenges in detecting them. We delve into the mechanics of deep fake creation, particularly the role of Generative Adversarial Networks (GANs), and discuss the real-world implications of this technology's misuse, from identity theft to political disinformation.

The presentation also examines cutting-edge detection methods, including AI-driven facial and audio analysis, and emerging technologies like block chain and quantum computing, which are being employed to verify the authenticity of digital content. Furthermore, we explore current applications of deep fake detection tools across social media platforms, law enforcement, and the media, while considering the ethical and legal issues surrounding the use and regulation of deep fake technology. By providing a comprehensive overview of the challenges and future trends in deep fake recognition, this presentation aims to equip audiences with the knowledge to navigate this evolving digital landscape.

Security Models and Strategies of Cloud Computing in Business and Public Administration

Zdzislaw Polkowski

WSG University, Poland

Globalization and the development of information and communication technologies (ICT) have caused significant changes in business and public administration in the European Union. Currently, the digital economy is driven by modern ICT systems that offer employers and employees new tools for effective operation. Cloud Computing systems deserve special attention, as they have an impact on business, public administration and the everyday life of citizens. Cloud Computing creates many benefits not only for users but also has a positive impact on the environment, which is in line with the & quot; European Green Deal & quot; strategy implemented in Europe. However, the security of Cloud Computing and many previously unseen threats are always in the focus of both cloud users and providers of this modern solution. Threats, such as zero-day exploits, advanced persistent threats (APTs), and insider attacks are some popular threats that strained the cloud services in business and public administration. These threats create a large barrier to widespread use and in many cases delay the further development of these systems. The regulations constantly introduced and updated in Europe take into account the development of secure Cloud Computing, sensibly regulating the ways of processing not only personal data, but also data processing in general. It is worth mentioning that research and analysis have been carried out in real companies and public administration institutions to build robust and secure cloud platforms that can tackle modern-day cyber threats. The analysis highlights that the best outcomes in cloud computing implementations are driven by the collaboration of interdisciplinary teams, specifically those composed of IT specialists, legal experts, and managers. This synergy ensures a holistic approach, balancing technical efficiency with legal compliance and strategic oversight. Special attention should be given to security strategies, such as robust data encryption, identity and access management, regular auditing, and compliance management to mitigate the negative consequences of security breaches in cloud environments. These advancements will be particularly beneficial for a wide audience, including IT professionals, researchers, students, and business leaders, all of whom play a role in implementing and managing cloud solutions. Furthermore, future research could explore how these interdisciplinary teams tackle emerging challenges, such as evolving regulatory landscapes, data sovereignty, and the integration of advanced technologies like AI and machine learning in cloud infrastructure. Such investigations would contribute to the ongoing development of best practices and position cloud computing as a more secure and sustainable solution in a rapidly changing digital ecosystem.

Keywords: Cloud Computing · IT security · Business · Public administration · European Green Deal

Role of Computing and Technology in Sustainable Development Goals

Murali Raman

Asia Pacific University of Malaysia

We are living in a highly Volatile, Uncertain, Complex and Ambiguous (VUCA) world, which threaten the long-term sustainability of global business, economy and natural living. The Sustainable Development Goals (SDGs) outlined by the United Nations serve as a global roadmap for achieving a more equitable and sustainable future. As the world grapples with pressing challenges such as climate change, poverty, and inequality, computing and technology have emerged as powerful tools for addressing these complex issues. This keynote explores the pivotal role of computing and technology in advancing the SDGs, highlighting specific applications and their potential impact.

Computing and technology can contribute to the SDGs in various ways. For instance, in the realm of climate change, advanced analytics can be used to optimize energy consumption and identify sustainable energy sources. Artificial intelligence can enhance the efficiency of renewable energy systems and predict natural disasters, enabling early warning systems and disaster preparedness. Additionally, computing-powered solutions can facilitate sustainable agriculture by optimizing crop yields, reducing water consumption, and minimizing the use of harmful chemicals.

In the context of poverty and inequality, computing and technology can empower marginalized communities through digital inclusion and education. Affordable and accessible internet connectivity can bridge the digital divide, providing opportunities for education, healthcare, and economic development. Furthermore, online platforms can facilitate access to financial services, enabling individuals and small businesses to escape poverty.

Computing and technology can play a crucial role in promoting good health and well-being. Telemedicine and remote healthcare services can improve access to quality healthcare, especially in rural and underserved areas. Additionally, data analytics can be used to identify disease outbreaks and track their spread, enabling effective public health interventions.

In summary, computing and technology offer immense potential for addressing the SDGs. By harnessing the power of these tools, we can create a more sustainable, equitable, and prosperous world for all. This keynote uses case examples from different parts of the world, showcasing how SDGs can be addressed with proper utilisation of computing and technology. However, realizing this potential requires collaboration among governments, businesses, and academia to develop innovative solutions and ensure equitable access to technology.

Cybersecurity in the Age of 4.0 IR

Srinath Doss

Botho University, Botswana

In the era of Industry 4.0, where cyber-physical systems, the Internet of Things (IoT), and artificial intelligence drive unprecedented connectivity and automation, cyber security has never been more critical. As industries leverage smart technologies to enhance productivity, efficiency, and innovation, the digital landscape expands, bringing both opportunities and vulnerabilities.

The convergence of operational technology with information technology in smart factories, autonomous vehicles, and interconnected supply chains creates a complex and dynamic environment. A single security breach can disrupt production, compromise sensitive data, and even endanger lives. Cyber attacks are evolving in sophistication, targeting the very backbone of our digital infrastructure.

Cybersecurity in the age of Industry 4.0 demands a proactive and multi-layered approach. It requires the integration of advanced threat detection systems, robust encryption protocols, and comprehensive risk management strategies. Organizations must foster a culture of cybersecurity awareness, ensuring that employees at all levels understand and adhere to security best practices.

Collaboration across sectors is essential. Governments, private enterprises, and academic institutions must work together to develop innovative solutions, share threat intelligence, and establish global cybersecurity standards. Continuous investment in research and development is crucial to staying ahead of cybercriminals and safeguarding our digital future. In this digital revolution, cybersecurity is not just a technical necessity but a fundamental pillar of trust and resilience. By embracing a holistic and forward-thinking approach, we can protect our assets, secure our data, and pave the way for a safer, more interconnected world.

Keywords: Industry 4.0 · Cyber security · Internet of things · Cyber Attack · Risk Management

Trust in the Realm of Generative AI – Opportunities and Threats

Arvind Benegal

Chief Risk Officer & Sr. Advisory Consultant, USA

Trust is of paramount importance in Information Technology systems and data confidentiality, integrity and availability. AI and Generative AI have reshaped the technology landscape. A balanced approach to innovation is needed that incorporates both Trust and Risk Management. The threats represent the clear and present risks and challenges posed by emerging technologies, while the opportunities speak to the benefits brought by these innovations. This highlights the dual nature of technological advancement. Wherever there is an opportunity, there is also a significant threat.

The opportunities include the efficiency gains of automation of repetitive tasks and enhanced decision-making. Personalization addresses tailored experiences for users in healthcare, finance, and customer service. Innovation allows for creation of new products, services, and business models. Democratization of technologies provides access, challenging elite layers, and Just in Time Knowledge suggests just enough to solve problems.

The risks include Bias and Fairness since AI systems may perpetuate or amplify existing biases. Security vulnerabilities pose increased attack surfaces and potential misuse of AI capabilities. Data Privacy presents risks of unauthorized data usage and breaches, and Prompt Injections increase exposure to sophisticated AI attacks.

The Audit Focus Areas should consider Data Integrity. It must be ensured that data used by AI systems is accurate, complete, and secure, and ethical AI establishes frameworks to prevent bias and ensure fairness. Accountability and transparency are achieved by creating mechanisms for auditing AI decisions and actions. Institutional guidance provides adherence to IT governance frameworks such as COBIT to manage AI risks. Robust control measures are implemented to enhance trust. For instance, ISO/IEC 42001 is an international standard that provides a governance framework for implementing and continually improving artificial intelligence management systems.

Trust is built by by incorporating strategies for IT Auditors. Continuous Monitoring regularly reviews AI systems and processes for compliance and performance. Collaboration with stakeholders by engaging with developers, data scientists, and business leaders aligns AI initiatives with ethical standards. Education and Awareness train employees on the risks and best practices related to AI. And finally, a Risk-Based Approach is adopted by prioritizing audits and controls based on the risk levels associated with different technologies.

Machine Learning and Data Analytics at Scale - Rise of Unified AI and Data Platforms

Kumar Gangwani

BDB.ai, India

The last 10+ years have seen a breakneck pace of innovation in areas of Data Analytics, Machine Learning and AI, something that remained unseen in many decades prior. The availability of inexpensive storage, affordable compute, innovations in parallel computing frameworks, global availability of open-source tools and libraries, automation of data engineering, algorithmic innovations and availability of specialized hardware like GPUs and TPUs have ushered in the modern Data and AI era.

The open-source ecosystem has democratized access to cutting-edge tools and technologies to the larger technology community. For instance, frameworks like TensorFlow and PyTorch allow developers and researchers to build sophisticated models quickly and efficiently.

Yet, most organizations, despite good intentions, sizeable investments in technology and large technology teams, struggle to scale after successfully operationalizing a few initial analytics and ML use cases. In fact, the rate of failures of corporate data science and machine learning teams is so high that the notion of the data and AI-driven enterprise often seems a myth.

Despite the availability of a plethora of tools for data engineering and machine learning, most organizations fail to build and scale with these effectively. Part of this has to do with the overpromise around capability of enterprise tools sold to them, which sometimes are good at a particular solution like ingestion or discovery only; this results in organizations purchasing a number of tools that don't integrate well with each other or scale for increasing data volumes and diverse data demands. But that's not the only reason. Fragmented, inconsistent and incomplete data is often a factor responsible for failure of Machine Learning projects. Integrating ML workflows into production environments, handling versioning, scaling models and automating monitoring – a discipline being referred to as ML Ops – remain challenging for many businesses; and add to that rising complexity of algorithms itself. Most technical resources have specialization in perhaps one or two areas, but a very limited view of the entire data chain, and struggle to solve platform scale issues.

The whole industry today longs for an integrated approach that allows operation of analytics and ML at scale. The past few years have seen the rise of a few unified platforms that integrate the entire data infrastructure and plethora of data processes – from ingestion (batch, micro batch, real-time), data cleansing, enrichment and preparation, storage into data lake, with right security and governance, notebook/tools to experiment with and deploy Data Science and ML models, enable ML automation, use of AI services and self-service dashboards and visualization – all under one roof. These unified platforms

greatly simplify and accelerate the time to commercialize analytics use cases, often 4–5 times faster than commonly used tools, and the end-to-end automation of all data processes greatly reduces resource effort by a large margin, providing for sizeable cost savings. This integrated capability offered by unified data and AI platforms is the future of enterprise analytics and Machine Learning.

Implementing Cyber Security Best Practices to Protect Your Crown Jewels

Ram Kumar G.

Global Automotive Company, India

In a hyper-connected digital world, cyber threats are ever growing and evolving, putting confidential data supporting IT infrastructure at risk. Cyber crimes are common today and no organization is immune to cyber attacks. The cyber crime industry has grown into a multi-billion dollar illegal business and the profile of the attacker has changed over the years – from script kiddies to well-organized hacker groups supported by nation states.

In this prevailing scenario, it is important for organizations to be cognizant of the cyber threats and proactively put in place cyber security measures by way of security controls, processes backed up by skilled talent. A strong understanding of the multi-faceted domains of cyber security is essential to appreciate the range and depth of the field. This exposure leads to exploring the security best practices that need to be implemented in organizations as per their risk profile to safeguard their critical IT infrastructure and confidential business information including intellectual property in what has come to be characterized as "crown jewels". The security controls are defined by the results of a risk assessment exercise, situational awareness of ongoing cyber attack campaigns enabled by inputs from cyber threat intelligence, obligations to customers and regulatory requirements for cyber security and data privacy.

Keywords: Cyber security · Best practices · Crown jewels · Cyber attacks · Data protection · Critical assets · Cyber threats · Security risks

Smart Spatial Computing and Its Application

P. Bhanu Prasad

Future Technology, France

Spatial computing is an emerging technology that blends digital and physical worlds, allowing users to interact with smart devices in more seamless and immersive ways. It leverages a range of technologies, including augmented reality (AR), extended reality (XR), Artificial intelligence, Internet of things (IoT), haptic feedback, computer vision, sensor fusion, geospatial data analysis, and spatial mapping. Spatial computing integrates various technologies to interpret the physical world and merge it with digital information, effectively combining both realms. These experiences can range from fully immersive environments to the addition of digital elements in the real world, creating interactive experiences that enhance our understanding of spatial relationships. At its core, spatial computing empowers users to visualize, manipulate, and analyze complex spatial data in real time with ease.

Advancements in hardware and software enable developers to create rich, immersive, interactive environments that captivate users. This evolution not only improves the overall user experience but also opens up new possibilities. Despite the significant potential of spatial computing, challenges persist. Concerns regarding data privacy, security, and ethical considerations need to be addressed before these technologies can be widely deployed. Additionally, the lack of standardization and interoperability among different spatial computing platforms presents further obstacles to widespread adoption.

This immersive, natural way of interacting can enrich our lives in numerous ways, enhancing hobbies like entertainment, gaming, and sports, as well as educational experiences and professional activities. The applications of spatial computing are extensive and diverse, impacting vital sectors that shape daily life, including navigation, astronomy, virtual home design, learning through gaming, skill development, healthcare and fitness, virtual shopping, industrial settings, remote assistance, architecture and design, the stock market, and logistics and supply chain management.

Spatial computing signifies a paradigm shift in how we interact with various environments, unlocking a vast array of new applications. Here are some key applications of spatial computing that can help those interested in pursuing a career in this field get started.

Keywords: Spatial computing · Extended reality (XR) · Industrial Internet of Things (IIoT) · Sensor fusion, Digital Twins · Artificial intelligence, Haptic

Enhancing Recommender Systems with Extensions of Fuzzy Sets for Considering Stakeholders' Uncertainty in Software Requirements Prioritization

Vassilis C. Gerogiannis

University of Thessaly, Larissa, Greece

In large, complex and distributed software development projects, it is often not feasible to implement all of the requirements in a single software release because of budget, time and staff constraints. Software requirements prioritization is an important activity in software requirements engineering which aims to evaluate and prioritize the candidate software features to be implemented in the next release of a software system. Requirements prioritization is often applied iteratively, according to various criteria, by the multiple involved project stakeholders who may have different roles, needs and knowledge. In large-scale software development projects, it is often practically impossible to ensure that all stakeholders have complete knowledge on all candidate requirements. Thus, stakeholders may show some degree of uncertainty and hesitation, as it is difficult to precisely evaluate each requirement according to each prioritization criterion. This talk will present results of our ongoing research work on combining Recommender Systems and soft computing techniques based on extensions of Fuzzy Sets to address problems of requirements prioritization in the context of large and complex software development/maintenance projects. Recommender Systems can be a useful solution to information overload situations when stakeholders have to evaluate a large number of alternatives. Fuzzy sets extensions, such as Intuitionistic Fuzzy Sets, can deal with stakeholders' uncertainty and hesitation regarding the prioritization criteria importance and requirements ratings. The results from applying our methods are promising since they demonstrate that they can be effectively used to support multiple stakeholders to prioritize large requirements sets.

Keywords: Software Requirements Prioritization · Recommender Systems · Intuitionistic Fuzzy Sets

Generative Adversarial Network (GAN): An Efficient Tool for Data Generation and Data Completion

Prabir Kumar Biswas

IIT, Kharagpur, India

Generative Adversarial Network (GAN) is a relatively recent concept in deep generative learning wherein two neural networks are pitted against each other in a zero-sum non-cooperative game to match a non-stationary distribution to an intractable stationary distribution. In this lecture we shall talk about how the generative modelling capability of GAN can be exploited to design efficient frameworks for image/video inpainting. Semantic inpainting refers to realistically filling up large holes in an image or video frame. Our approach is to first train a generative model to map a latent noise distribution to a natural image manifold and, during inference time, search for the best-matching noise vector to reconstruct the signal. The primary drawback of this approach is its inference time iterative optimization and lack of photo-realism at higher resolution. In this talk both of the abovementioned shortcomings are addressed. This is made possible with a nearest neighbour search-based initialization (instead of random initialization) of the core iterative optimization involved in the framework. The concept is extended for videos by temporal reuse of solution vectors. Significant speedups of about 4.5-5x on images and 80x on videos is achieved. Simultaneously, the method achieves better spatial and temporal reconstruction qualities.

Next, we will pose the following question and try to get the answer - 'Do we at all need an iterative inference framework?' - to answer this, a data-driven parametric network is trained to directly predict a matching prior for a given masked image. This converts an iterative paradigm to a single feed-forward inference pipeline with around $800\times$ speedup. Finally, recent advancements in high-resolution GAN training are leveraged to scale an inpainting network for higher resolution.

Artificial Intelligence and Machine Learning: From Early Symbolic Systems to More Complex System

Salisu Mamman Abdulrahman

Kano University of Science and Technology, Nigeria

Artificial Intelligence (AI) and Machine Learning (ML) have evolved significantly over the years, moving from simple symbolic systems to more complex and sophisticated systems capable of learning and adapting. Advances in AI and ML systems have revolutionized various sectors, changing how people interact with technology. For four decades (mid-1950s until the mid-1990s), Symbolic AI was the dominant paradigm. These algorithms work by processing symbols that represent objects or concepts in the real world and their relationships. In Symbolic AI the main approach is to use logic-based programming where rules and axioms are used to perform inference and deduction. High expectations surrounding AI's potential to solve complex problems and mimic human intelligence were met with disappointing results, leading to what is known as the AI Winter. Despite the challenges faced by researchers during the AI Winter, the field experienced a revival in the 1990s and early 2000s. AI Researchers shifted their focus from Symbolic AI and rule-based systems to statistical and probabilistic approaches that are more promising. The emergence of machine learning algorithms, such as Support Vector Machines and Bayesian networks, enabled AI systems to demonstrate tangible successes in various applications. Machine learning systems have evolved dramatically over the years, progressing from basic models capable of simple tasks to complex neural networks that drive cutting-edge AI applications. Throughout this evolution, AI and ML have transitioned from rule-based systems to learning-based approaches, greatly expanding their capabilities and applications. In this keynote address, we will present an overview of their journey from inception to recent advances.

Keywords: Artificial Intelligence · Machine Learning · Symbolic System · Complex System

Attention Mechanism for Abnormal Behavior Detection from Video Surveillance

Vinothina V.

Kristu Jayanti College (Autonomous), Bengaluru, India

In todays fast-paced environment, detecting abnormal behaviour in real time is increasingly challenging, yet video monitoring remains vital for public safety. Traditional surveillance systems still rely heavily on manual observation, which is not only time-consuming but also prone to errors, especially when managing large volumes of footage. While machine learning (ML) algorithms have been widely used for behaviour detection, they often struggle with the complexity of unstructured video data. These algorithms typically require vast amounts of labelled data and can falter when dealing with imbalanced datasets, reducing their accuracy in identifying rare, yet critical, behaviours. Moreover, ML techniques tend to have slower learning rates, limiting their effectiveness in real-time scenarios and large-scale data applications. In contrast, deep learning models, especially convolutional neural networks (CNNs), have proven more effective by automatically learning feature representations directly from raw video data. However, CNNs still face challenges when it comes to capturing the temporal dynamics of behaviour across video frames. To address these challenges, the proposed method integrates a hybrid model that combines the strengths of CNNs and transformers with a context-based attention mechanism.

Modelled after the human brain's ability to focus on the most pertinent aspects of a scene, attention mechanisms integrated into deep learning frameworks have significantly improved both the accuracy and efficiency of abnormal behaviour detection. When combined with CNNs, these attention mechanisms emphasize key spatio-temporal features. Spatial attention allows the model to detect critical body movements and gestures, while temporal attention tracks changes in behaviour over time, identifying sudden or repetitive actions. This approach allows for real-time tracking of individual actions within the video frame, enabling timely alerts and interventions when abnormal behaviours are detected. When evaluated against existing methods, the proposed solution demonstrated superior performance, effectively overcoming the limitations associated with traditional machine learning approaches in behaviour recognition.

Keywords: Attention Mechanism · Abnormal Behaviour Detection · Video Surveillance · Deep Learning · Spatial Attention · Temporal Attention · Contextual Information

Zero Trust and Its Impact

Chander Shekhar Sharma

AUNZ - Cybersecurity Unit, Australia

The Zero Trust security model represents a fundamental shift in how organizations approach cybersecurity. Unlike traditional security frameworks that rely on a strong perimeter to protect internal networks, Zero Trust operates on the principle of "never trust, always verify." This approach assumes that threats can exist both inside and outside the network, and therefore, every access request—whether from a remote employee or an internal user—must be authenticated and authorized. By focusing on the identity of users and devices rather than the network location, Zero Trust transforms the way organizations secure their resources. At the heart of Zero Trust is the concept of identity as the new security perimeter. This means that organizations must implement robust identity verification mechanisms to ensure that only authorized individuals can access sensitive information. Multi-factor authentication (MFA) is a critical component of this strategy, requiring users to provide multiple forms of identification before gaining access. Additionally, risk-based adaptive authentication can further enhance security by adjusting access levels based on various factors, such as the user's location, device, and behaviour patterns. This continuous verification process helps organizations mitigate the risk of unauthorized access and data breaches.

The benefits of adopting a Zero Trust identity strategy are substantial. For instance, organizations can significantly reduce the likelihood of data breaches by ensuring that every access request is scrutinized. This not only protects sensitive information but also fosters a culture of security awareness among employees. Furthermore, Zero Trust can improve the user experience by streamlining access to resources while maintaining stringent security measures. For example, remote workers can securely access company applications from any location without compromising security, thereby enabling greater flexibility and productivity.

Use cases for Zero Trust identity are increasingly relevant in today's digital landscape. Organizations that have adopted cloud services or remote work policies can leverage Zero Trust principles to secure their environments effectively. For instance, a financial institution might implement Zero Trust to protect customer data by requiring MFA for all transactions and continuously monitoring user behaviour for anomalies. Similarly, a healthcare provider could use Zero Trust to ensure that only authorized personnel can access patient records, thereby complying with regulations and safeguarding sensitive information. By humanizing the approach to cybersecurity, Zero Trust not only addresses technical challenges but also aligns with the evolving needs of organizations and their employees in a rapidly changing digital world.

Empowering Academic Advising Through Smart Computing and Predictive Analysis

Ahmed Al-Brashdi

UTAS, Oman

The role of academic advising has evolved from static and transactional to proactive support, emphasizing innovative approaches to address the variations of students' performance and needs. By leveraging historical data on student performance, course complexity, and former evaluation, predictive analytics can identify students at risk of academic difficulty. An early warning system with comprehensive student data and predictive analysis based on historical data enables timely interventions, such as academic coaching or tutoring, to improve student retention and success. Moreover, using predictive analytics and AI will optimize resource allocation by balancing advisor workloads, and optimizing course scheduling. This data-driven approach contributes to cost-efficiency and improved student satisfaction. This approach can personalize the academic experience for each individual student by recommending tailored academic plans based on individual student strengths, goals, and performance patterns without overloading advisors. This approach fosters student engagement and accelerates progress towards degree completion. In UTAS – Oman, an in-house system enables advisors and students to effectively and efficiently engage with personalized study plans and course scheduling with emphasis on on-probation students. It shows improvements in student satisfaction and resource allocation.

Keywords: Academic advising · Predictive analysis · Smart computing · AI

Best Practices of Research Ethics

Vijay Singh Rathore

Shree K. Karni Universe College, Jaipur

Research ethics constitute a cornerstone of scientific inquiry, ensuring the credibility, integrity, and societal relevance of scholarly endeavours. This study delineates the fundamental principles governing ethical research practices, emphasizing adherence to honesty, objectivity, and transparency in all stages of research, from data collection to dissemination. Key ethical mandates include the avoidance of fabrication, falsification, and plagiarism, as well as the active protection of confidentiality, intellectual property rights, and the dignity of research participants.

The role of supervisors is highlighted as pivotal in fostering an ethical research environment, ensuring adherence to legal and institutional norms, and mentoring researchers in responsible practices. Equally, researchers are tasked with diligent participation in their projects, rigorous compliance with safety protocols, and acknowledgment of collaborative contributions.

Ethical challenges, such as authorship disputes, data ownership, and conflicts of interest, are addressed through structured protocols and mutual respect among collaborators. Furthermore, responsible dissemination of findings underscores the importance of accuracy and societal impact, avoiding premature or misleading claims. This framework underscores the collective responsibility of the research community to uphold ethical standards, thereby advancing scientific progress while maintaining public trust.

Contents – Part IV

Cyber Security

Automation of Endpoint Security Using Machine Learning 3
Simongi Patel and Vipul Chudasama

Comparative Analysis of Vulnerability and Threats Related to Blogging Platforms 11
Nidhi Khare and Sachin Kadam

Cybersecurity Threats to Autonomous Vehicles: A Deep Learning Approach 25
V. Sathyavathy

Advanced Threat Intelligence Systems for Early Warning of Cyber Hacking . . . 35
T. Sharath and A. Muthukumaravel

Reconciliation - Backdoor Access Finding Strategies with Legacy Applications 50
Bhushan B. Chavan, Harsh Soni, Lakhichand Khushal Patil, and Kalpesh A. Popat

Enhancing Cyber Hacking Breach Classification Through Integrated Preprocessing and Feature Extraction Techniques 67
J. Christina Deva Kirubai and S. Silvia Priscila

An Ensemble Method for Insider Threat Detection Based on User Activities Analysis Using Bi-LSTM and GA Optimization 80
A. Anju, Adline Freeda, Krithika Venket, Babymol Kurian, D. Sudha, and W. Mercy

Systemization of Knowledge: Privacy Preservation for Decentralized Non-financial Applications 88
Purnima Ahirao and Deepti Reddy

Identifying Application-Level Distributed Denial-of-Service (DDoS) Attacks with Machine Learning (ML) 105
Muhammad Atif Yaqub, Muzzamil Mustafa, Saad Hussain Chuhan, Basit Sattar, Muhammad Zulkifl Hasan, Muhammad Zunnurain Hussain, Muhammad Saad Qaiser, Zain ul Abdin, Afshan Bilal, and Jibran Ali

"Cybersecurity Threats and Trust Dynamics in Digital Payment Systems: An Analysis of Domestic Fraud and User Perspectives" 116
Vishal Ramaiya, Pramod Goyal, and Neeraj Kumar Dubey

Enhanced Web Browser Forensics: Innovative Methodologies for Evidence Collection and Analysis 139
Nishant Navinbhai Joshi and Sunil Lalchand Bajeja

DDOS Attack Detection and Prevention Using MIEC Method in SDN Environment 165
R. Karthikeyani and E. Karthikeyan

A Lossless Image, Video and Audio Encryption Method Based Hybrid Chaotic Map and DNA Encryption 176
Vrushali Khaladkar and Manish Kumar

Enhancing Cloud Security for Structured Data: An AES-GCM Based Format-Preserving Encryption Approach 196
S. Vidhya

Hacktivism or Cyber Warfare? Decoding the Motivations Behind Cyber Attacks Targeting Israel 206
Niraj Kumar Singh and Sunil Bajeje

Designing a Secure E-Voting System with Ganache: Challenges and Solutions 233
Garima Mathur, Pushpraj Singh Chauhan, and Privi Dubey

Computer Application for Sustainability

Consumer Usage Intention Factors for Mobile Food- Delivery Applications Using Task Technology Fit Approach 247
Shweta Shirolkar and Kanchan Patil

Optimizing Query Execution Time in Relational Databases Using Cost-Based and Index Merge Strategies: A Case Study on E-Commerce Data 264
Patel Nilen and Himansu Maniar

Securing Sign-Up and Sign-In in the Indian Market: Fraud Prevention in CIAM for CRM - A Comprehensive Review and Future Directions 274
Nitesh Behare, Manesh R. Palav, Sharmishthadevi B. Deshmukh, Shrikant Waghulkar, and Vinod Pal

MERN Stack Website for Government Officials 302
Rohit A. Patil, Kalpana S. Pawase, Jyoti Sarwade, Kaustubh Chaudhari, and Tushar Suryawanshi

Brain Tumor Diagnosis using Deep Learning: A Systematic Review and Meta-analysis of MRI Image-Based Studies 309
Patel Kruti Dineshbhai and Himanshu Maniar

A Study on Employee Engagement Factors of Information and Communication Technology (ICT) Workforce 320
Gowri Ramakrishnan and Neeraj Kumar Dubey

A Simulated Manufacturing Study on Rope Size and Number of Splits on the Productivity While Applying the Simplified Drum-Buffer-Rope Methodology 338
Jagannathan Sekkizhar and Nagarajan Vivek

Effect of Technology Acceptance Among Employees in the Second Half of Their Working Life 348
Firdaus Bashir

A Study on the Effects of Social Media Platforms on Graduate Students' Academic Achievement 363
Jatinkumar B. Kotadiya and Amit K. Patel

Enhancement of Forest Fire Assessment by KDCPMNN Approach in Sikkim, India Using Remote Sensing Images 379
Kapila Sharma and Gopal Thapa

Decentralized Payment Gateway: Bridging DeFi with Metamask 395
Ashish Revar, Khooshi Sonkar, and Kalpesh Wandra

Impact of Social Media Marketing in the Cosmetic Industry to Cross the Chasm in India 409
Narendra Rathnaraj, C. Rajasree, S. Amrutha, and Arul Rajan

Ubiquitous Computing of Low-Cost Pollutant Sensors for Monitoring Pollutant Gases in the Construction Sector 425
Rebakah Geddam, Mohd Zuhair, Kush Shah, Karm Vyas, and Naitik Kanani

Envisioning the AI-Driven Future: A Delphi Study on the Transformative Role of Artificial Intelligence in Higher Education 435
Meeta Joshi and Janvi Bhindi

IoT Integration for Smart Agriculture: Prioritizing Key Factors Using WASPAS Approach 452
M. K. P. Naik, Anand Jaiswal, Ashish Yadav, and Teena Singh

Transforming EHR Systems Using the HyperLedger Fabric Blockchain Framework 464
Neha A. Samsir and Arpit A. Jain

Author Index 477

Cyber Security

Automation of Endpoint Security Using Machine Learning

Simongi Patel and Vipul Chudasama(✉)

Nirma University, Ahmedabad, India
{23MCC017,vipul.chudasama}@nirmauni.ac.in

Abstract. With the increasing usage of the internet and technology, cyber threats are also increasing which target endpoint devices. This work focuses on the importance of securing the endpoints and the highlights the role of Intrusion Detection Systems (IDS) in mitigation of cyber threats. This research work provides insight into various threats related to the Network environment and analysis of intrusions. To mitigate such threats, we propose customized IDS mechanisms for secured endpoints using different ML techniques with the NIDS dataset. The patterns of anomaly are identified which leads to proactive threat detection. By continuously updating predefined rules based on emerging threats, automated Endpoint Security mechanisms can be deployed to mitigate security threats effectively. The results suggest some models perform well to detect threats.

Keywords: Endpoint security · Machine learning · Automation · IDS

1 Introduction

With the increase in usage of internet and technology, the cyber-attacks are increasing rapidly. The complexity of cyberattacks is increasing each passing day. An intentional attempt carried out by attackers to gain the access of individual's or organization's system is called cyber-attack. Attackers look for the system vulnerabilities to carry out network-based attacks on it [13].

Cyber attackers injects viruses like malware on endpoints and obtain malicious access to any system [4]. During the year 2021, the ratio of worldwide internet users who were attacked with atleast 1 malware-class, containing mostly ransomware were around 15.45%. As of 2023, over 72% of business organizations were affected by ransomware attacks. It is very crucial for organizations to detect APTs at correct time and circulate warnings so that whole APT attack can be prevented on the systems [14].

Cybersecurity is very critical in order to ensure access only to authorized users by seeking potential measures to protect assets like information, networks and data against threats. Cyber-security professionals protect networks, servers, intranets, and computer systems [16]. However, according to the [15], the signature-based tools are also unable to detect the situations for malicious activities of hackers. This is the limitation which highlights the evolving nature of cyber threats and need for more effective detection methods.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 3–10, 2025.
https://doi.org/10.1007/978-3-031-86305-9_1

Endpoint detection and response (EDR) tools are the advanced cybersecurity technology which integrates solutions of endpoint antivirus and endpoint management to identify, examine, and remove malicious software which results into exploitation of system. They are used with to larger security systems such as security information and event management (SIEM), vulnerability management, and incident response tools. By the year of 2029, the market of EDR is going to witness huge growth [18].

2 Need for Automation in Endpoint Security

Due to the evolving landscape of cyberthreats, it becomes necessary to protect endpoint devices and implementing automation of endpoint security. By adoption various ML techniques, the paper aims to enhance the threat detection and response. For achieving endpoint security, adopting an automated approach is best because it eliminates repetitive tasks and leads to faster incident response to security threats. It gives rapid threat detection and response without human intervention and automatically applies patches across all endpoints thus reducing the attack surface.

3 Literature Survey

Authors [1] highlight the limitations of traditional security tools, which often fail to detect advanced breaches, causing disruptions and data theft. Fragmented security infrastructures exacerbate visibility gaps, increasing the risk to organizations.

In [2], explores the application of machine learning (ML) techniques for cybersecurity, focusing on ensemble learning algorithms to improve the detection of cyberattacks and address data issues such as cross-validation and model performance.

Authors [3] emphasize the importance of endpoint security in protecting enterprise networks, particularly in light of the growing use of remote work and personal devices. The study highlights common attack vectors and defense mechanisms.

[4] reviews anomaly detection techniques such as Local Outlier Factor (LOF) and autoencoders for identifying unknown intrusions. The authors also discuss the challenges posed by malware and fileless attacks that bypass traditional security systems.

[5] examines security in distributed environments, such as cloud computing and IoT, focusing on the integration of Security Content Automation Protocol (SCAP) and Trusted Computing Group (TCG) for continuous security auditing.

[6] analyzes the role of machine learning in improving cybersecurity by detecting anomalies. It addresses challenges related to noisy and imbalanced data and highlights proactive defense strategies using intelligent systems.

[7] discusses network softwarization and automation for complex environments, proposing models for automating security and traffic flow configuration.

[8] explores the use of dynamic honeypots for gathering threat intelligence and improving network security. The study focuses on the challenges associated with static honeypots and the benefits of using dynamic deployment.

[9] addresses the difficulty of detecting insider threats and compares traditional Endpoint Protection Platforms (EPP) with more advanced Endpoint Detection and Response (EDR) systems.

[10] highlights the integration of AI and ML in threat detection to prevent ransomware and zero-day attacks. The authors review the use of the EMBER dataset to train models for detecting executable files and assess the effectiveness of anomaly detection in cybersecurity.

4 Methodology

This research paper outlines the steps for automating endpoint operations to enhance security. The approach includes (Fig. 1):

1. System Interaction: Engages both legitimate user actions and potential attacker behaviors, such as unauthorized access and exploitation of vulnerabilities. This helps identify normal and anomalous behaviors, enabling proactive threat detection and response.
2. Data Pre-processing: Collects and preprocesses data from sources like user activity logs, network traffic, and system events for analysis.
3. Machine Learning Techniques: Applies ML models, including Random Forest, KNN, and SVM, to pre-processed data to detect patterns and anomalies, improving threat detection and reducing false positives.
4. Intrusion Detection System (IDS): Analyzes ML output to detect anomalies and trigger alerts for potential threats.
5. Rule Engine: Integrates with IDS to apply and update predefined rules for threat identification and response. Automated mechanisms mitigate threats by blocking suspicious traffic and updating rules based on new threats.

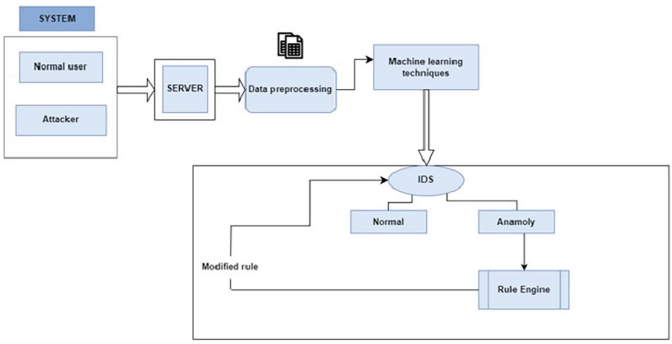


Fig. 1. Proposed Methodology

4.1 Sample IDS Rules

Rule 1: alert any any -> any any (msg:“Unusual system interaction detected”; sid:1001;)
This rule is triggered when the IDS detects an unusual system interaction.

Rule 2: alert any any -> any any (msg:“Preprocessed data anomalies detected”; sid:1002;)

This rule detects anomalies in pre-processed data. Rule 3: alert any any -> any any (msg:“Anomalies detected in trained ML models”; sid:1003;)

This rule detects anomalies in trained machine learning (ML) models. Rule 4: alert any any -> any any (msg:“Intrusion detection system alert”; sid:1004;)

This rule is triggered to indicate an alert from the intrusion detection system itself. Rule 5: alert any any -> any any (msg:“Automated response to intrusion”; sid:1005;)

This rule is triggered when the IDS initiates an automated response to an intrusion or security event.

4.2 Proposed Algorithm

Algorithm 1 System Automation Algorithm

```
procedure main
    interactions ← system_interaction()
    pre-processed_data ← data_pre-processing(interactions)
    trained_models ← apply_ml_techniques(pre-processed_data) with hyperparameters tuning/
    anomalies ← intrusion_detection_system(trained_models)
    responses ← rule_engine(anomalies)
    updated_rules ← update_rules(anomalies, responses)
    return updated_rules
end procedure
```

5 Results and Discussion

5.1 Dataset Description and Feature Selection

The “Network Intrusion Detection” dataset [19] contains simulated network traffic from a U.S. Air Force LAN, with 25,192 training and 22,544 test observations. Each observation has 41 attributes (3 qualitative and 38 quantitative), labeled as “Normal” or “Anomalous” for intrusion detection. Feature selection identifies the most relevant features for the ML model to enhance performance and reduce costs. Figure 3 depicts feature selection method with the chi-squared metric (χ^2), which measures feature-target independence by comparing observed and expected frequencies. The top 10 features are selected based on their chi-squared scores for model training and testing (Fig. 2).

5.2 Performance Evaluation of ML Models

To evaluate IDS performance, we use the following metrics:

– True Positive (TP): Correctly predicted positive instances.
– False Positive (FP): Incorrectly predicted positive instances. – True Negative (TN): Correctly predicted negative instances.

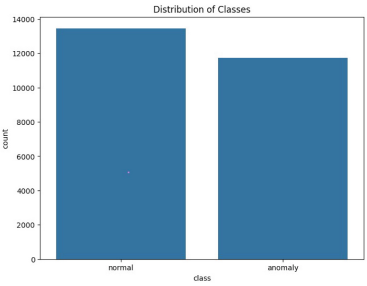


Fig. 2. Class Distribution

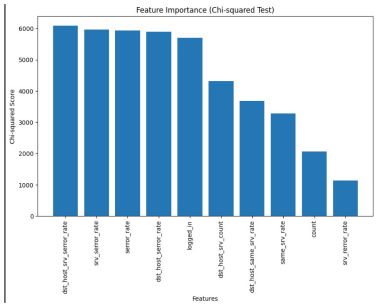


Fig. 3. Feature Selection

– False Negative (FN): Incorrectly predicted negative instances.
– Accuracy: Measures the overall correctness of the predictions.

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

– Precision: Measures the proportion of correctly identified positive cases out of all cases predicted as positive.

$$\text{Precision} = \frac{TP}{TP + FP}$$

– Recall: Measures the proportion of actual positive cases that were correctly identified by the IDS.

$$\text{Recall} = \frac{TP}{TP + FN}$$

– F1-score: A combined metric that balances precision and recall, providing a single measure of model performance.

$$\text{F1} - \text{score} = \frac{2 * Precision * Recall}{Precision + Recall}$$

5.3 Performance Evaluation of Various Ensemble ML Models

Ensemble learning combines multiple models to enhance performance. Here, we employed three ensemble techniques:

1. Voting Classifier: A Voting Classifier is an Ensemble Learning Method which takes the combined predictions of multiple base ML models. In this approach I have used Gradient Boost, KNN and Random Forest.
2. Bagging Classifier: A bagging classifier makes a new model from independent weak models like trees, which are trained in parallel. In this approach, I have used a base model as Decision Tree.
3. Stacking Classifier: A stacking classifier combines multiple base models by using a meta-learner to train the outputs of the base models. In this approach, I have used Random Forest, KNN, and Decision Tree as base models and Logistic Regression as a meta-model.

Table 1. Performance Metrics of ML Models

Model	Accuracy	Precision	Recall	F1-score	TP	TN	FP	FN
RF	0.9543	0.9330	0.9846	0.9581	2633	2176	189	41
KNN	0.9618	0.9634	0.9848	0.9612	2580	2267	98	94
SVM	0.9575	0.9456	0.9760	0.9601	2610	2215	150	64
DT	0.9632	0.9683	0.9622	0.9652	2573	2281	84	101
NB	0.9104	0.8800	0.9626	0.9194	2574	2014	351	100
LR	0.9220	0.9074	0.9498	0.9281	2540	2106	259	134
GB	0.9682	0.9676	0.9727	0.9701	2601	2278	87	73

Table 2. Performance Metrics of Ensemble Models

Model	Accuracy	Precision	Recall	F1-score	TP	TN	FP	FN
Voting	0.9746	0.9642	0.9775	0.9708	2614	2268	97	60
Bagging	0.9701	0.9687	0.9745	0.9716	2606	2281	84	68
Stacking	0.9682	0.9647	0.9778	0.9712	2612	2267	98	62

6 Conclusion and Future Scope

In conclusion, we observe that in Table 1, the accuracy of Gradient Boosting is the highest at 96.82%, despite a slightly higher false positive rate of 87 compared to the Decision Tree's false positive rate of 84. However, in Table 2, we can observe that the Bagging Classifier achieves the highest accuracy of 96.98% with a false positive rate of 84, matching the false positive rate of the Decision Tree. Thus, by observing both accuracy and false positive rate of ML models and Ensemble ML models, we can conclude that the Bagging Classifier is more suitable as it achieves highest accuracy and also maintains a balanced false positive rate. To extend this research and identify future opportunities, one can use hybrid model which combines different techniques can be useful. Active learning and explainability techniques will make it more trustworthy for endpoint security analysts.

References

1. Arfeen, A., Ahmed, S., Khan, M.A., Jafri, S.F.A.: Endpoint detection & response: a malware identification solution. In: Proceedings of the 2021 International Conference on Cyber Warfare and Security (ICCWS), p. 18 (2021). https://doi.org/10.1109/ICCWS53234.2021.9703010
2. Mehta, S., Darji, M., Joshi, H.: Threat prediction using ensemble learning algorithm to provide end-point security. In: Proceedings of the 2022 International Conference on Electronics and Renewable Systems (ICEARS), pp. 1427–1434 (2022). https://doi.org/10.1109/ICEARS53579.2022.9751963

3. Kamruzzaman, A., Ismat, S., Brickley, J.C., Liu, A., Thakur, K.: A comprehensive review of endpoint security: threats and defences. In: Proceedings of the 2022 International Conference on Cyber Warfare and Security (ICCWS), pp. 1–7 (2022). https://doi.org/10.1109/ICCWS56285.2022.9998470
4. Kim, S., Hwang, C., Lee, T.: Anomaly based unknown intrusion detection in endpoint environments. Electronics **9**(6), 1022 (2020). https://www.mdpi.com/2079-9292/9/6/1022. https://doi.org/10.3390/electronics9061022
5. Aslam, M., Gehrmann, C., Björkman, M.: Continuous security evaluation and auditing of remote platforms by combining trusted computing and security automation techniques. In: Proceedings of the 6th International Conference on Security of Information and Networks, pp. 136–143. Association for Computing Machinery (2013). https://doi.org/10.1145/2523514.2523537
6. Sarker, I.H.: Machine learning for intelligent data analysis and automation in cybersecurity: current and future prospects. Ann. Data Sci. **10**(6), 1473–1498 (2022)
7. Kaur, H., Tiwari, R.: Endpoint detection and response using machine learning. J. Phys. Conf. Ser. **2062**(1), 012013 (2021). https://doi.org/10.1088/1742-6596/2062/1/012013
8. Bussa, S., Sisto, R., Valenza, F.: Security automation using traffic flow modeling. In: Proceedings of the 2022 IEEE 8th International Conference on Network Softwarization (NetSoft), pp. 486–491 (2022). https://doi.org/10.1109/NetSoft54395.2022.9844025
9. Bartwal, U., Mukhopadhyay, S., Negi, R., Shukla, S.: Security orchestration, automation, and response engine for deployment of behavioural honeypots. In: Proceedings of the 2022 IEEE Conference on Dependable and Secure Computing (DSC), pp. 1–8 (2022). https://doi.org/10.1109/DSC54232.2022.9888808
10. Chandel, S., Yu, S., Tang, Y., Zhou, Z., Huang, Y.: Endpoint protection: measuring the effectiveness of remediation technologies and methodologies for insider threat. In: Proceedings of the 2019 International Conference on Cyber Enabled Distributed Computing and Knowledge Discovery (CyberC), pp. 81–89 (2019). https://doi.org/10.1109/CyberC.2019.00023
11. Rathod, V., Parekh, C., Dholariya, D.: AI & ML based anomaly detection and response using ember dataset. In: Proceedings of the 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 1–5 (2021). https://doi.org/10.1109/ICRITO51393.2021.9596451
12. Aslan, Ö., Aktuğ, S.S., Ozkan-Okay, M., Yilmaz, A.A., Akin, E.: A comprehensive review of cyber security vulnerabilities, threats, attacks, and solutions. Electronics **12**(6), 1333 (2023)
13. Sudar, K.M., Deepalakshmi, P., Nagaraj, P., Muneeswaran, V.: Analysis of cyberattacks and its detection mechanisms. In: Proceedings of the 2020 Fifth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 12–16 (2020). https://doi.org/10.1109/ICRCICN50933.2020.9296178
14. Hoque, N., Bhuyan, M.H., Baishya, R.C., Bhattacharyya, D.K., Kalita, J.K.: Network attacks: taxonomy, tools and systems. J. Netw. Comput. Appl. **40**, 307–324 (2014). https://doi.org/10.1016/j.jnca.2013.08.001. https://www.sciencedirect.com/science/article/pii/S1084804513001756
15. Abdullahi, H.B. (n.d.). Developing intelligent cyber threat detection systems through advanced data analytics
16. Li, Y., Liu, Q.: A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments. Energy Reports **7**, 8176–8186 (2021). https://doi.org/10.1016/j.egyr.2021.08.126. https://www.sciencedirect.com/science/article/pii/S2352484721007289
17. Kumar, A., Singh, J.: Unveiling the anatomy of phishing attacks: strategies, targets, and cybersecurity solutions. In: Proceedings of the 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT), pp. 1–6 (2023). https://doi.org/10.1109/EASCT59475.2023.10393225

18. Sjarif, N.N.A., et al.: Endpoint detection and response: why use machine learning? In: Proceedings of the 2019 International Conference on Information and Communication Technology Convergence (ICTC), pp. 283–288 (2019). https://doi.org/10.1109/ICTC46691.2019.8939836
19. Sampadab17, Network Intrusion Detection Dataset (2021). https://www.kaggle.com/datasets/sampadab17/network-intrusion-detection. Accessed 31 July 2024

Comparative Analysis of Vulnerability and Threats Related to Blogging Platforms

Nidhi Khare[1(✉)] and Sachin Kadam[2]

[1] Bharati Vidyapeeth (Deemed to be University), Pune, India
nidhikhare89@gmail.com
[2] IMED, Bharati Vidyapeeth (Deemed to be University), Pune, India
sachin.a.kadam@bharatividyapeeth.edu.in

Abstract. With the proliferation of technology and its advancements, blogs play a key role in writing content on various topics replacing the traditional ways of maintaining diaries. Blogging is becoming a passion for many people due to the services provided by free blogging platforms; bloggers are making a realm of it. Blogging platforms are widely used today by individuals, businesses, and organizations for content creation, marketing, and sharing information. However, they are also vulnerable to security threats, which can lead to data breaches, website defacements, or other cyber-attacks. In this paper, we have studied different blogging platforms like WordPress, Blogspot and Wix for analysing, the associated security concerns, and the threats and vulnerabilities these platforms face. These mentioned blogging platforms are widely used by bloggers. The popularity of these platforms is due to several unique features such as ease of its usage, flexibility, extensive feature sets. Here, security is the prime concern for bloggers as the content which are posted on these platforms are vulnerable to various threats and vulnerabilities. In this research paper, we performed the comparative analysis of vulnerability and threats related to three blogging platforms- WordPress, Blogspot and Wix by following the research methodology of experiment. The experiment was conducted using penetration testing on the above-specified blogging platforms to understand the better security features provided by them. Based on the vulnerabilities and threats identified for these platforms, it can be concluded that Blogspot platform provides better security features with less vulnerabilities and threats.

Keywords: Security · Blogging Platforms · Blogs · Blog Analysis · Vulnerabilities · Threats

1 Introduction

Blogging has experienced tremendous growth since the advancement of the internet and technologies. The traditional way of managing diaries has evolved into a diverse ecosystem of writing blogs based on various topics, that is, from personal experiences and hobbies to niche interests or professional expertise. A Blog is a popular way of

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 11–24, 2025.
https://doi.org/10.1007/978-3-031-86305-9_2

deliberation or sharing information published on the informational website which consists of discrete, often informal just like the style of diary text entries (posts). A typical blog combines textual content, images, videos and links to other types of blogs, pages of website, and other contents related to its concept. Platforms of Blogging are foundations for bloggers which allows them to create, organise and publish their contents on these platforms. These platforms provide tools and templates for publishing content, managing its appearance, and controlling the type of user interactions. The contents of blogs are published in reverse chronological order. These platforms allow the user to publish the post anytime, and displays the date and time when the blog is posted. The blogging platforms provide free or paid services for bloggers to publish and host their blogs [1]. Regular blogging can help people to find business online and help bloggers to make money from blogging. Some of the most popular blogging platforms include WordPress, Blogspot and Wix.

1.1 Threats and Vulnerabilities Related to Blogging Platforms

Threats and Vulnerabilities related to blogging platforms are weaknesses or flaws in these platforms that can harness the security of posting blogs on these platforms.

Vulnerabilities: A flaw (weakness) or a loophole in the system which exposes it to attack is called vulnerability. These vulnerabilities are related to one or more types of security threats. Some of the common vulnerabilities include Security misconfiguration, Injection attacks, Cross Site scripting, Vulnerable and outdated components, etc. [2]. Security vulnerability can distress five types of system securities that include: Accuracy, Privacy, Completeness, Serviceability, and unquestionability.

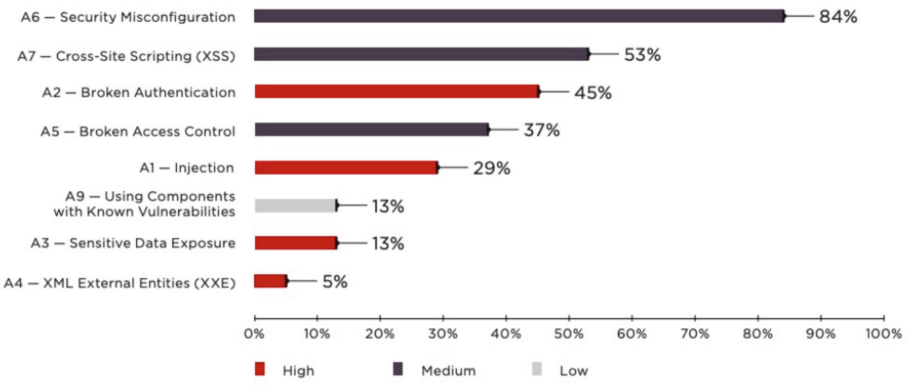


Fig. 1. Most common OWASP Top 10 vulnerabilities (percentage of web applications)

The OWASP's (Open Worldwide Application Security Project) most common top 10 vulnerabilities are shown in Fig. 1, it can be observed that the most common vulnerability is Security Misconfiguration.

Threats: A threat tries to exploit the vulnerabilities (loopholes) to gain unauthorized access to data or systems. It can also be considered as a malicious event or adverse effect which negative event which exploits the vulnerability. Natural, Human made, Accidental or Intentional are some of the threats [3]. Some of the common threats include data breach, personal attacks, identity theft, email phishing, tampering of data etc.

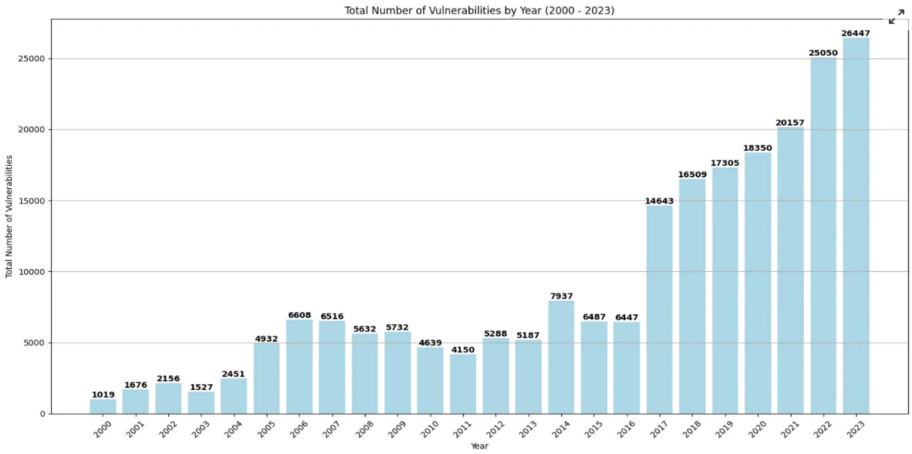


Fig. 2: Total number of Vulnerabilities by year (2000–2023) Source: Qualys.com

The number of vulnerabilities which are exploiting various web platforms are specified in Fig. 2, the total number of vulnerabilities by year (2000–2023) has been significantly increasing as shown in the graph.

Objectives of the Study

1) To study the vulnerabilities and threats related to blogging platforms of WordPress, Blogspot and Wix.
2) To compare the vulnerabilities and threats related to the above specified blogging platforms and find the platform which provides better security features

2 Research Methodology

Experimental research methodology has been applied on the three most popular blogging platforms-Wordpress, Blogspot and Wix. These platforms have been chosen due to combination of various unique features provided by them making it most popular. The researcher has created his own blogs on these blogging platforms and then applied Penetration testing tool to detect the vulnerabilities and threats or risks related to these blogging platforms. Pen-test free tool is being used to check the security vulnerabilities and threats in these three blogging platforms. This tool gives the complete report of threats and risks, their impact and category of vulnerabilities under which it comes.

The info, low, medium, and high risk or threat levels and vulnerabilities have been estimated using this tool which are as per OWASP listings of vulnerabilities and threats/risks [4]. The comparative analysis of threats and vulnerabilities has been performed from the result of pen testing to analyse the blogging platform with better security features (Fig. 3).

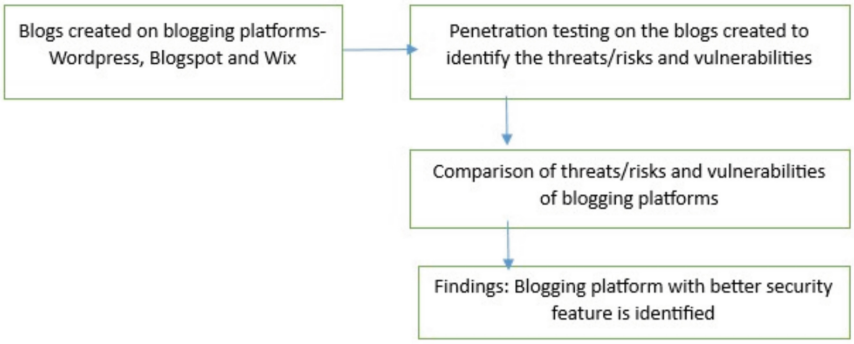


Fig. 3. System flow of Research

2.1 Experiment on WordPress, Blogspot and Wix Blogging Platforms

To make security comparison easier, the Blogs are created in all these platforms with almost similar types of contents. These three blogging platforms are:

a) WordPress
b) Blogspot
c) Wix

Pen-test free tool is being used to check the security vulnerabilities and threats in these three blogging platforms. This tool gives the complete report of threats/risks, their impact and category of vulnerabilities under which it comes.

a) WordPress

The detail report is given below (Table 1):

Detection of injection points which are unique	1
Spidered URLs	1
HTTP Requests number	9
Response received in average time	132 ms

Table 1. Risks/threats and Vulnerabilities related to WordPress blogging platform

S.No	Risks	Evidence	Risk Description	Vulnerabilities
1	Insecure cookie setting: missing Secure flag	Set-Cookie: country_code = GB; Path = /	The risk lies that attacker can get access which is unauthorised for the web session of victim	Security Misconfiguration
2	Insecure cookie setting: missing HttpOnly flag	Set-Cookie header(s) is responded by the server that does not specify the HttpOnly flag: Set-Cookie: country_code = GB	Cookie can be accessed by the attacker injecting Javascript code which is malicious, which can be sent to other site, leading to session hijacking	Security Misconfiguration
3	Missing security header: X-Content-Type-Options	X-Content-Type-Options HTTP security header are not included in Response headers	Phishing in browsers or Cross-Site scripting could be the possible attacks in this case	Security Misconfiguration
4	Missing security header: Content-Security-Policy	HTTP Content-Security-Policy security header or meta tag is not specified in the response	Target application becomes vulnerable to XSS, which attackers can easily exploit	Security Misconfiguration

(*continued*)

Table 1. (*continued*)

S.No	Risks	Evidence	Risk Description	Vulnerabilities
5	Robots.txt file found	–	Addition of endpoints cannot be contemplated as security measure as this file can be read and directly accessed by anyone	Security Misconfiguration
6	Server software and technology found	–	This information can be used to set up attacks which are specific against the software version and type	Security Misconfiguration

b) Blogspot

The detail report of pen testing of the blog created on Blogspot is follows (Table 2)-

Detection of injection points which are unique	55
Spidered URLs	6
HTTP Requests number	14
Response received in average time	284 ms

Table 2. Risks/threats and Vulnerabilities related to Blogspot blogging platform

S.No	Risks	Evidence	Risk Description	Vulnerabilities
1	Missing security header: Referrer-Policy	Referrer-Policy HTTP security header as well as the tag with name 'referrer' is not present in the response headers	If a web page is visited and user clicks on that link, the URL with full originating will be sent by browser in the Referrer header	Security Misconfiguration
2	Missing security header: Strict-Transport-Security	HTTP Strict-Transport-Security header Request / Response are not included in Response headers	It allows an attacker to enforce a proper HTTP connection to the server. Chances of extracting sensitive information and snooping on the network traffic	Security Misconfiguration
3	Missing security header: X-Content-Type-Options	HTTP Content-Security-Policy security header or meta tag are not included in Request/Response	XSS could be the possible attacks in this case	Security Misconfiguration
4	Missing security header: X-Content-Type-Options	X-Content-Type-Options HTTP security header Request / Response is not included	Phishing in browsers or Cross-Site scripting could be the possible attacks in this case	Security Misconfiguration

(*continued*)

Table 2. (*continued*)

S.No	Risks	Evidence	Risk Description	Vulnerabilities
5	Robots.txt file found	–	Addition of endpoints cannot be contemplated as security measure as this file can be read and directly accessed by anyone	Security Misconfiguration
6	Server software and technology found	–	This information can be used to set up attacks which are specific against the software version and type	Security Misconfiguration
7	Security.txt file is missing	–	It is important to have this file as it provides a specific channel for specifying security issues and vulnerabilities	Security Misconfiguration

c) Wix

The detail report is given below (Table 3):

Detection of injection points which are unique	2
Spidered URLs	2
HTTP Requests number	10
Response received in average time	176 ms

Table 3. Risks/threats and Vulnerabilities related to Wix blogging platform

S.No	Risks	Evidence	Risk Description	Vulnerabilities
1	Insecure cookie setting: missing HttpOnly flag	Set-Cookie header(s) that does not denote the HttpOnly flag	Cookies can be accessed by the attacker injecting Javascript code which is malicious, which can be sent to other site, leading to session hijacking	Security Misconfiguration
2	Insecure cookie setting: domain too loose	Set-Cookie:.wix.com	Cookies such as example.com can be sent with the sent requests. These cookies can be accessed and main site can be accessed via that site	Security Misconfiguration
3	Missing security header: Referrer-Policy	Referrer-Policy HTTP security header as well as the tag with name 'referrer' is not available in the response	If a web page is visited and user clicks on that link, the browser will specify the complete originating URL in the Referrer header, header of Referrer Policy is not set is assumed	Security Misconfiguration

(*continued*)

Table 3. (*continued*)

S.No	Risks	Evidence	Risk Description	Vulnerabilities
4	Unsafe security header: Content-Security-Policy	HTTP Content-Security-Policy security header with the security issues are included in Response header	An attacker can exploit this by executing arbitrary JavaScript code in the This can be exploited by an attacker to execute arbitrary JavaScript code in the framework of vulnerable application	Security Misconfiguration
5	Robots.txt file found	–	Addition of endpoints cannot be contemplated as security measure as this file can be read and directly accessed by anyone	Security Misconfiguration
6	Missing security header: Content-Security-Policy	HTTP Content-Security-Policy security header or meta tag is not specified in the response	Target application becomes vulnerable to XSS, which attackers can easily exploit	Security Misconfiguration

(*continued*)

Table 3. (*continued*)

S.No	Risks	Evidence	Risk Description	Vulnerabilities
7	Server software and technology found	–	This information can be used to set up attacks which are specific against the software version and type	Security Misconfiguration
8	Security.txt file is missing	–	It is important to have this file as it provides a specific channel for specifying security issues and vulnerabilities	Security Misconfiguration

2.2 Result and Analysis of Experiment

The comparative analysis of pen tests performed on blogging platforms-WordPress, Blogspot and Wix has been specified in Table 4.

Table 4. Comparative Analysis of Test Performed on Blogging Platforms

Test No	Test Performed	Wordpress	Blogspot	Wix
		Rank Rating (High/ Medium/ Low/ Info)	Rank Rating (High/ Medium/ Low/ Info)	Rank Rating (High/ Medium/ Low/ Info)
1	Missing Secure flag	Medium	Low	Info
2	Missing HttpOnly flag	Medium	Info	Medium
3	X-Content-Type-Options	Low	Low	Info
4	Content-Security-Policy	Low	Low	Low
5	Robots.txt file found	Low	Low	Low
6	Server software and technology	Low	Low	Low
7	Website accessibility	Info	Info	Info

(*continued*)

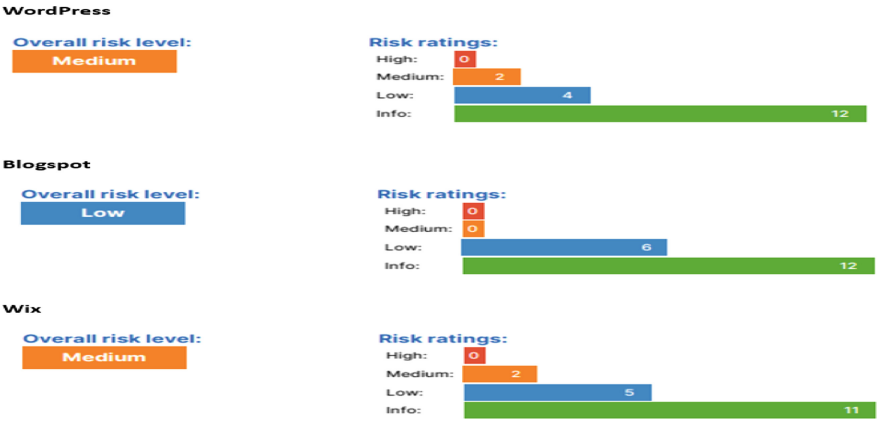
Table 4. (*continued*)

Test No	Test Performed	Wordpress	Blogspot	Wix
		Rank Rating (High/ Medium/ Low/ Info)	Rank Rating (High/ Medium/ Low/ Info)	Rank Rating (High/ Medium/ Low/ Info)
8	Vulnerabilities of server-side software	Info	Info	Info
9	Client access policies	Info	Info	Info
10	Absence of the security.txt file	Info	Info	Info
11	Untrusted certificates	Info	Info	Info
12	Enabled HTTP debug	Info	Info	Info
13	Secure communication	Info	Info	Info
14	Directory listing	Info	Info	Info
15	Strict-Transport-Security	Info	Low	Info
16	HTTP header – Referrer	Info	Info	Low
17	Set for cookies	Info	Info	Medium
18	HTTP header Content Security Policy	Info	Info	Low

The summary of comparative analysis report for blogging platforms- WordPress, Blogspot and Wix has been specified in Table 5. The ranking of risks/threats have been labelled as High, Medium, Low and Info based on the result shown on the Pen testing report**.**

Table 5. Comparative Analysis Summary Report

Blogging Platform	Test Rating				Overall Risk
	High	Medium	Low	Info	
Wordpress	0	2	4	12	Medium
Blogspot	0	0	6	12	Low
Wix	0	2	5	11	Medium



The comparison summary of no. of risks/threats based on the risk rating has been shown for three blogging platforms which has been studied in the paper. Thus, it is imperative from the above analysis that Blogspot has low and info types of risks/threats as compared to WordPress and Wix.

3 Conclusion

This research paper analyses the various security features related to security of WordPress, Blogspot and Wix blogging platforms. It focuses on different types of threats and vulnerabilities associated with blogging platforms. It can be concluded from the experiment carried out that out of all three blogging platforms, Blogspot provides better security features. The analysis report lists out less vulnerabilities and threats for Blogspot which have low impact as compared to Wix and WordPress after conducting the experiment through Penetration testing. The main vulnerability which has been identified through this experiment is 'Security Misconfiguration' which exposes most of the threats. This vulnerability needs to be fixed as this may exploit the security features of the blogging platforms listed above. Therefore, the conclusion can be drawn that better security is being provided with Blogspot as compared to other two blogging platforms. It is hence an important finding from the perspective of bloggers who are constantly being using the various blogging platforms for self-expression.

References

1. Vaishnavi, V.R.: Unveiling the Blogosphere: Exploring the Evolution, Impact, and Future of Blogging
2. Nidhi, K., Dr. Sachin, K.: Review of threats and vulnerabilities associated with blogs. Shodh Sarita J. **7**(26), 131–137 (2020)

3. Nidhi, K., Sachin, K.: A systematic review of blogging: perspectives of security challenges and opportunities. Indian J. Tech. Educ. **47**(2) (2024)
4. owasp.org

Cybersecurity Threats to Autonomous Vehicles: A Deep Learning Approach

V. Sathyavathy(✉)

Department of Computer Technology, KG College of Arts and Science, Coimbatore, India
Sathyavathy.v@kgcas.com

Abstract. In the upcoming years, autonomous cars are predicted to completely transform transportation by offering greater accessibility, efficiency, and safety. Autonomous vehicles are a critical component of smart mobility, which is a fundamental component of smart cities. Significant new Cybersecurity vulnerabilities are also introduced by the intricate sensor systems, onboard computers, and vehicle-to-infrastructure connectivity needed for self-driving cars. However, human safety and quality of life may be negatively impacted by vulnerabilities in Autonomous vehicles. Automotive Cybersecurity is getting a modulation point. The attacks of cyber incidents have grown drastically, threatening safety and carrying functioning important implications. It is more crucial to protect automobiles, mobility apps, and Internet of Things gadgets from the cyberattacks of self-driving cars in a proactive manner. The motivation of threat actors is changing to focus on having a significant and large-scale control on connected cars and motion assets. Deep Learning is a key element of smart cities, and it is closely coupled with autonomous attacks and defences.

Keywords: Cybersecurity · IoT · Artificial Intelligence · Cyberattack · Autonomous Vehicles

1 Introduction

Compared to typical cars, autonomous vehicles offer far larger attack surfaces due to their millions of lines of software code, several electronic control units, and multiple data flows. If a cyberattack is successful, it could endanger driver privacy by gaining access to onboard sensor data or jeopardise passenger safety by interfering with essential vehicle systems. The intricacy of autonomous car systems also complicates the application of conventional security measures like anti-virus software and patching [1]. Car manufacturers use encryption and access controls to handle these issues. They also use intrusion detection systems and anomaly detection techniques to spot threats. By closely observing vulnerability testing, redundancy, compartmentalization, and other recommended procedures, self-driving car safety and dependability gains can be achieved while simultaneously shielding cars from cyberattacks. As this revolutionary technology gains traction over the next ten years, more research and development centered on Cybersecurity for autonomous vehicles will be essential.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 25–34, 2025.
https://doi.org/10.1007/978-3-031-86305-9_3

2 Importance of Automotive Cybersecurity

Protecting critical car systems against cyberattacks, such the brakes and steering, requires automotive Cybersecurity. In actuality, this safety feature is essential for protecting not just the motorist from possible injury but also other road users and pedestrians.

Vehicles are gathering and transmitting sensitive data, including position data in real time, as they become more and more like computers on wheels. In order to protect this data and guarantee driver's privacy and security, vehicle cybersecurity is essential.

Maintaining trust in the cybersecurity of connected cars is critical given their increasing ubiquity. This entails securing not just the car but also the user's private data, social security numbers, address, numbers of credit card (Fig. 1)

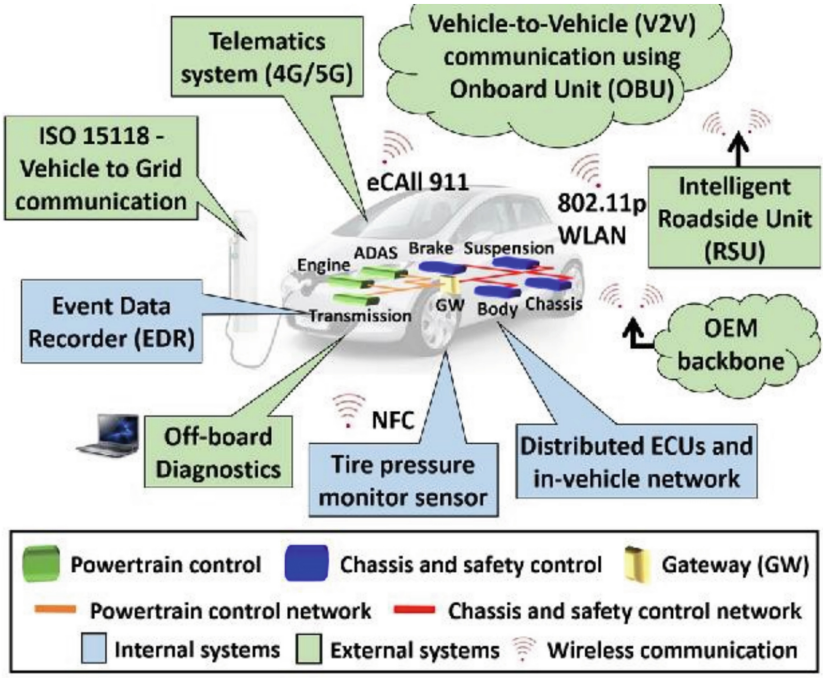


Fig. 1. Overview of Autonomous Vehicle

3 IoT Attacks

IoT (Internet of Things) attacks on vehicles of autonomous are increasing alarm due to the increasing connectivity and complexity of these systems (Fig. 2)

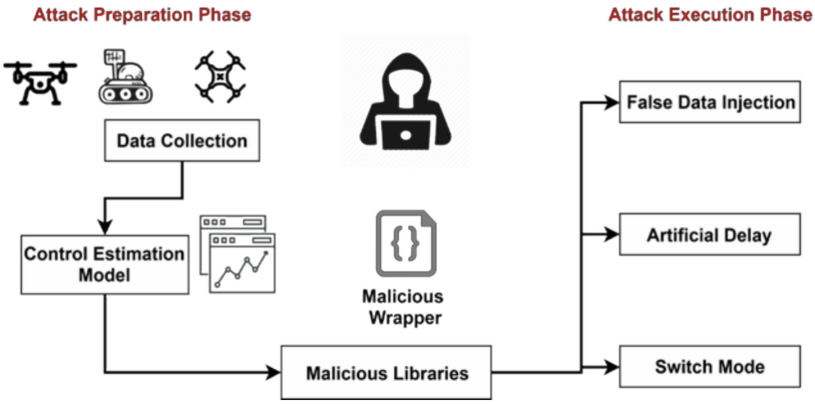


Fig. 2. Autonomous Systems and IoT Security

4 Solutions for Cyberattacks in Vehicles

4.1 Use of Encryption

In order to secure data transfer inside the car as well as during connectivity with external networks, sophisticated algorithms are used. For instance, encryption can protect data transferred between a car's manufacturer's servers and the car itself, guaranteeing the privacy of sensitive data like user preferences and position.

4.2 Firewalls

These serve as a barrier, keeping an eye on, and controlling the flow of data into and out of the car's network. By preventing unwanted access, firewalls help guarantee that only authorized communication.

4.3 Systems for Detecting and Preventing Intrusions (IDPS)

These technologies monitor car networks in real time, quickly identifying and stopping any questionable activity. For example, an IDPS might spot an odd request to access the vehicle's control system and stop it right away, averting possible damage.

4.4 Verification of Firmware and SecureBoot

This guarantees that the software on vehicle control units is only validated and authorised. In this way, it stops dangerous software from getting installed, much as how a smartphone only permits the download and use of verified apps.

4.5 Control of Access

Strong authentication procedures must be put in place to stop unwanted control. This may entail multifactor authentication or biometric verification to guarantee that the vehicle's systems are only accessible by authorised users or the legitimate owner.

4.6 Updates via OTA Over-the-Air (OTA)

Security upgrades are necessary for the software in a car to operate properly. Encrypted data transmission and stringent verification procedures are used in secure over-the- air updates to make sure the changes are real and unaltered.

4.7 Penetration Testing and Security Audits

System vulnerability evaluations are similar to standard physical examinations. These tests and audits assist in locating such vulnerabilities and stop attackers from taking advantage of them.

4.8 Training and Awareness on Cybersecurity

Teaching business experts best practices is like preparing warriors for combat. It gives them the information and abilities needed to recognise cybersecurity threats and take appropriate action.

5 IoT Security Threats

5.1 Usage of Default Passwords

Most companies do not reveal the default passwords on the devices they ship. For example, this frequently occurs with light controller systems, household routers, and safety cameras. The fact that default passwords are easily compromised by hackers due to their widespread knowledge poses one of the largest threats to Internet of Things security [2].

5.2 Insecure Communication

IoT safety problems rise from the point that many IoT devices transmit information over system of networks without encrypting them. The best method to confirm a protected connection is to use protocols like Transport Layer Security (TLS) and transport encryption. Devices separated over many networks ensure a safe and distant connection while maintaining the confidentiality of transferred data

5.3 Leaks of Private Information

Even with just the knowledge of IP (Internet Protocol) addresses from unpatched IoT devices, skilled data thieves might cause a great deal of damage. It is possible to ascertain a user's exact location and residential address using these addresses. In order to safeguard the Internet of Things connection and conceal your IP address, several experts in internet security suggest utilizing a virtual private network, or VPN.

5.4 Automation and Artificial Intelligence (AI)

AI technologies are already widely applied. However, automation has a disadvantage in that the AI network and the infrastructure can be brought down by a single programming error or bad algorithm [3] (Fig. 3).

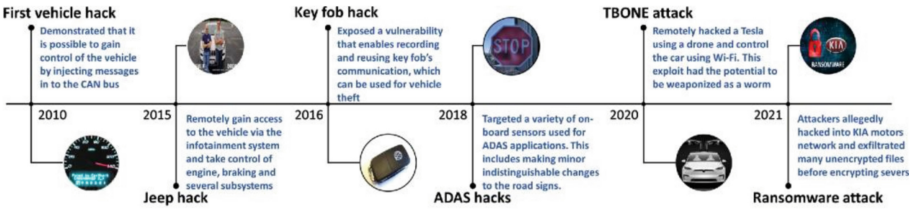


Fig. 3. Timeline of Cybersecurity Attack

6 Cybersecurity of Autonomous Vehicles

With more electrical components inside the automobile, there are more attack vectors that an attacker can try to use to try and gain access to the information and data inside the vehicle or even change how it operates. Automobile manufacturers want to produce AVs that follow a security cycle of design and development to lessen cybersecurity issues [4]. Designing, creating, and keeping an eye on software and hardware over the span of the vehicle's life is all part of this cycle. To address cybersecurity issues, certain manufacturers have taken a variety of measures, including assessing threat intelligence, sharing the information with external partners, developing and delivering security patches for cars.

In order to reduce cybersecurity concerns, automakers work to create AVs that adhere to a security cycle of design and development. This cycle includes designing, producing, and monitoring hardware and software over the course of the vehicle's lifetime. Some automakers have implemented a wide range of strategies to address cybersecurity concerns, such as analysing threat intelligence, disseminating the information to outside partners, and creating and distributing security patches for auto- mobiles. In order to combat the increasingly complicated cyber-threat situation (such as manipulation, fraud, and interference with safely operated vehicles), comprehensive cybersecurity solutions are needed

6.1 Automotive Cybersecurity Standards and Regulations

Because an antivirus programme makes decisions based on a variety of facts and artificial intelligence, Cybersecurity and system failure have emerged as the most important factors in deciding safety and security. The core of the automotive supply chain is cybersecurity [5]. The objective of this standard is to encourage original equipment manufacturers (OEMs), component, software suppliers, and mobility services of vehicles to implement more robust cybersecurity protections.

Potential Benefits of Autonomous Vehicles. The advancement of autonomous technology holds the potential to transform the automotive sector and alter the modes of transportation [6]. Wide-ranging societal advantages from autonomous cars could include higher road safety, more effective transit, and greater access to mobility. One of the main benefits of autonomous vehicles is increased road safety. Nowadays, human error such as driving while distracted or intoxicated, speeding, or making poor decisions for more than 90% of auto accidents.

Furthermore, by adhering to traffic laws and maintaining appropriate speeds and distances between vehicles, autonomous cars are anticipated to promote safer driving. In order to get a full 360-degree view of their surroundings, onboard computers can combine data from many sensors and cameras. In addition to almost completely eliminating blind spots, this may identify dangers, bicycles, and pedestrians more accurately than human drivers. Autonomous vehicles can cooperate and coordinate to prevent crashes and maintain efficient traffic flow thanks to linked vehicle-to-vehicle (V2V) communication protocols.

New Cybersecurity Risks Posed by Vehicle Autonomy and Connectivity. Although connected and driverless cars have a lot of potential advantages, they also present serious new cybersecurity risks and vulnerabilities. Although linked and driverless cars provide a plethora of potential advantages, they also raise serious privacy and cybersecurity dangers and erode public trust in autonomous technology [7]. Up to 150 electronic control units (ECUs) with 100 million lines of code can be found in autonomous cars. Compared to conventional vehicles, this significant increase in complexity results in a greater number of potential weaknesses that attackers could take advantage of. Crucial automotive operations like steering, braking, and acceleration are controlled by ECUs.

Innovative cybersecurity solutions for autonomous vehicles (AVs) involve advanced technologies and approaches to address the unique challenges posed by these complex systems.

Machine Learning and AI. There are three types of Anomalies Detection in Cybersecurity.

Abnormality Detection. Use AI and ML to detect anomalies in vehicle behavior, network traffic, and sensor data, identifying potential threats in real-time. *Predictive Maintenance.* Employ predictive analytics to anticipate and address cybersecurity vulnerabilities before they are abused.

Adaptive Safety. Implement adaptive security structures that dynamically implement security measures created on the current threat landscape.

Importance of Proactive Cybersecurity Measures. Significant cybersecurity lapses that result in crashes and casualties might substantially endanger public acceptance of autonomous driving technology as it advances. People lose control of the vehicle, thus they need assurances that these cars can safely traverse the digital world. Proactive cyber defences are therefore crucial for preventing attacks and maintaining consumer confidence [8]. Manufacturers need to take quick action in the event of cyber incidents. Because modern cars have about 100 million lines of code, vulnerabilities are inevitable. Software bugs, unpatched components, and supply chain issues will all present potential for exploitation. In addition to external hackers, opponents might come from nation-states or even company employees.

Challenges Implementing in Cyberthreats in Autonomous Vehicles. To implement all preventive measures to prevent cyberthreats to the automotive industry. There are many challenges to make it practically possible.

Human error. Unintentionally increasing hazards can happen with employees, independent contractors, and outside vendors. Cybersecurity awareness and education are

vital to prevent unethical hacking and hazards in order to reduce the risk of insider attacks [9].

High Volume of Data. Autonomous vehicles generate a lot of data on the occupants and their environment. Utilizing and managing the massive amount of data for alerts and insights can be somewhat challenging at times. Ensuring that privacy regulations be observed and preventing unauthorized access to this data are growing concerns.

Measures to Avoid Future Cyberthreats in Autonomous Automotives. Future Cyberdangers in the autonomous cars sector must be avoided through the proactive and comprehensive strategy.

Security by Design. It means including Cybersecurity concerns from the very beginning of the design and development of autonomous vehicles. Throughout the whole lifecycle of the car, security is implemented from the conception of maintenance *Regular Software Updates.* Ensurity of security patches and updates applied to all autonomous vehicles software on regular basis. Over the Air (OTA) mechanism is updated that is reliable and safe to securely send updates

AI-Based Multi-Layer Security. To safeguard the vehicles network and data, use many layers of security systems, including firewalls, IDS (Intrusion Detection Systems) and encryption

7 Proposed Algorithm for CAV

CNN-LSTM algorithm combines temporal pattern recognition with LSTMs and spatial feature extraction with CNNs, making it perfect for identifying complex cyberattacks with both spatial and temporal properties. Ensemble methods are used for combining multiple models to improve robustness and accuracy in detecting cyberattacks. Graph Neural Networks (GNN) is better in identifying the connections and interdependencies between various parts of a car's network, helpful in spotting organized assaults. The deep learning algorithm is used to detect CAN attacks [10].

The long-term knowledge dependency (LTK) method as a time-recurrent neural network (RNN). LSTM's flow is similar to that of RNN. The manner that cells function in the case of LSTM differs from that of RNN approaches. The input, candidate, forget, and output gates make up each LSTM unit. Data is categorized by the forget gate according to whether they should be saved or discarded. The outcome gate always regulates the concealed state in the LSTM, whereas the input gate refreshes the cells. Furthermore, LSTM solves the disappearance and implosion-gradient problems in the RNN learning process by including an embedded memory block and gate structure (Table 1 and Fig. 4)

Table 1. Comparison Results of Deep Learning Algorithm

Models	Labels	Precision (%)	Recall(%)	F1-Score(%)	Accuracy(%)
CNN	Two	77	87	81	86
CNN-LSTM	Six	91	94	92	94.01
CNN-LSTM	Three	98	97	97	97.44

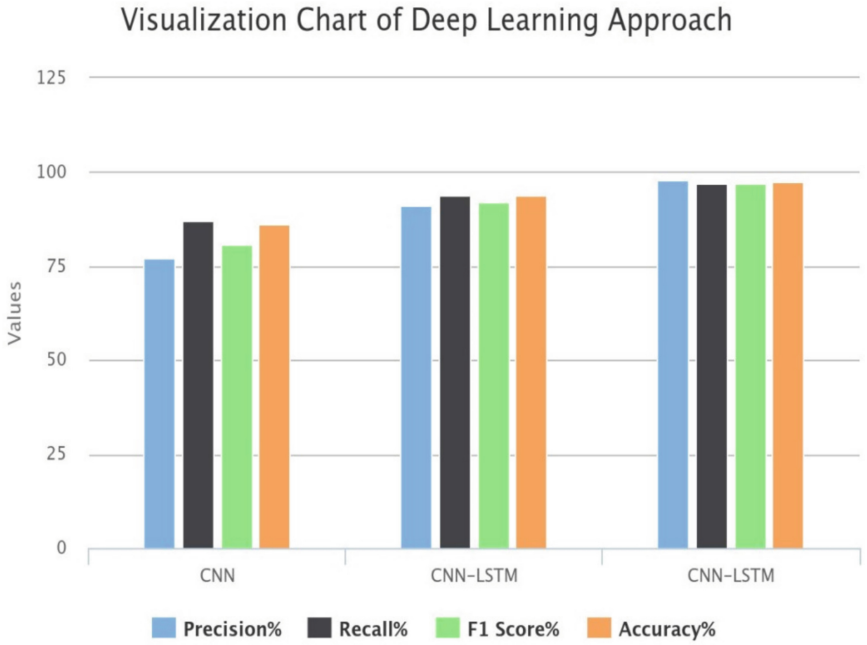


Fig. 4. Visualization of Deep Learning Approach

8 Evaluation Metrics

The conventional metrics of accuracy, recall, precision, and F1-score were employed in order to assess the proposed system. The indications of confusion metrics, which include true-positive (TP), false-positive (FP), true-negative, and false negative (FN), are used to construct the assessment metrics (Fig. 5).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{FP} + \text{FN} + \text{TP} + \text{TN}} \times 100\%$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100\%$$

$$\text{F1} - \text{score} = 2 * \frac{\text{precision} \times \text{sensitivity}}{\text{precision} + \text{sensitivity}} \times 100\%$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \times 100\%$$

Fig. 5. Evaluation Metrics

9 Results

The deep learning approach is used as the proposed method to detect the various types of attacks from the vehicle network. An actual network with fuzzing, spoofing, replaying, and regular traffic was used to test the system. Thirty percent of the datasets were put aside for testing, while seventy percent were used for training. The conventional methods for identifying attack messages in a CAN bus are inapplicable (Table 2 and Fig. 6).

Table 2. Performance Accuracy of Models against the proposed system

Models	Accuracy %	Attack Types
Deep Learning Model	95%	Normal and Attacks
Deep Learning Model	85%	DoS,Command,Injection,Malware Attacks
Generative Adversial Networks	95%	DoS, Fuzzing and Gear attacks
LSTM	80%	Spoofing, Replay and Flooding attacks
Machine Learning	90%	DoS, Fuzzing, spoofing attacks
Neural Network- LSTM	90%	DoS, Fuzzing, spoofing attacks
CNN-LSTM	97.5%	Dos, Fuzzing, spoofing ,Replaying

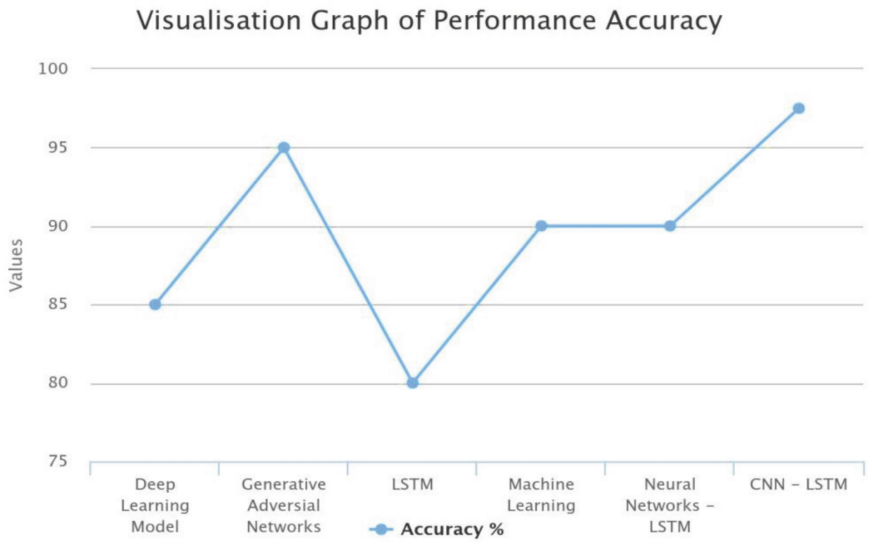


Fig. 6. Visualization Graph of Performance Accuracy

10 Conclusions

The Internet of Things and the fast growth of the car industry have made the network of autonomous vehicles more sophisticated and well-established. The self-driving car connects to entertainment or satellite navigation systems, offering numerous amenities. However, because the automatic intelligent vehicle network is connected to the Internet for remote access, autonomous vehicles that offer these neural features run the risk of being attacked remotely. Traffic activity is broadcast domain in nature. It has not been easy to create a security system that works. Consequently, the intrusion detection system based on artificial intelligence models has offered solutions to combat the increased risk associated with vehicle networks. They can also be expanded to include different security system designs for safe data processing inside the intricate networks of autonomous cars. The suggested systems received an accuracy score of 97.5% overall. When compared to current systems, these empirical results performed better. In the future, cutting-edge artificial intelligence can be used to further enhance our system.

References

1. Guizzo, E.: How Google's Self-Driving Car Works. IEEE Spectrum Online. https://spectrum.ieee.org/howgoogle-self-driving-car-works. Accessed 23 Jul 2023
2. Sharma, B., Sharma, L., Lal, C., Roy, S.: Anomaly based network intrusion detection for IoT attacks using deep learning technique. Comput. Electr. Eng. **107**, 108626 (2023)
3. Baggenstoss, P.M.: New restricted boltzmann machines and deep belief networks for audio classification. In: 14th ITG Conference, Speech Communication, pp. 1–5 (2021)
4. Zhu, Y., Yan, W.Q.: Traffic sign recognition based on deep learning. Multimedia Tools Appl. **81**, 17779–17791 (2022)
5. Autonomous Vehicle Technology and Its Impact on Transportation in 2023 - Sky Is Not The Limit. Sky Is Not the Limit (2023). https://gurmeetweb.com/autonomous-vehicle-technology-and-itsimpact-on-transportation-in-2023/
6. Frąckiewicz, M.: Cyber-physical Security for Transportation and Mobility. TS2 SPACE (2023). https://ts2.space/en/cyber-physical-security-for-transportation-and-mobility
7. NHTSA. Automated Vehicles for Safety. www.nhtsa.gov/technology-innovation/automated-vehicles-safety. Accessed 23 Jul 2023
8. Bordoloi, U., Chakraborty, S. Jochim, M., Joshi, P., Raghuraman, A., Ramesh, S.: Autonomy-driven emerging directions in software-defined vehicles. In: 2023 Design, Automation & Test in Europe Conference & Exhibition (DATE), pp. 1–6 (2023)
9. Khastgir, S., et al.: Distributed ODD awareness for connected and automated driving. Transp. Res. Procedia **72**, 3118–3125 (2023)
10. Mokander, J., Floridi, L.: Operationalizing AI governance through ethics-based auditing: an industry case study. AI and Ethics **3**(2), 451–468 (2023)
11. Marie, L.: NVIDIA Enters Production with DRIVE Orin, Announces BYD and Lucid Group as New EV Customers, Unveils Next-Gen DRIVE Hyperion AV Platform," Published by Nvidia Newsroom, Press Release, 2022. Accessed 26 Jan 2024
12. Girdhar, M., Hong, J., Lee, H., Song, T.-J.: Hidden markov models-based anomaly correlations for the cyber-physical security of EV charging stations. IEEE Trans. Smart Grid **13**(5), 3903–3914 (2022)

Advanced Threat Intelligence Systems for Early Warning of Cyber Hacking

T. Sharath[1(✉)] and A. Muthukumaravel[2]

[1] Department of Computer Science, Bharath Institute of Higher Education and Research (BIHER), Chennai, India
chakram.babu@gmail.com

[2] Faculty of Arts and Science, Bharath Institute of Higher Education and Research (BIHER), Chennai, India
dean.arts@bharathuniv.ac.in

Abstract. As we continue to deploy state-of-the-art cybersecurity measures, targeted cyber-attacks have increased in scale and frequency. A Research on a New Advanced Threat Intelligence System for Early Warning in Cyber Attack based on Realtime Monitoring, Machine Learning and Big data Analytics The system combines, tidies and synthesizes data across disparate sources. It uses supervised and unsupervised machine learning models to assess the risks in that case. ATIS have introduced big data tools to conduct fast large-scale data analysis to predict threats online and activate alerts; In a controlled evaluation of ATIS, there was seen to be an improvement in the accuracy and reduction in false positives associated with threat detection. This new method gives an end-to-end cybersecurity solution for proactive security services, which allows organizations to gather such required data on a timely basis and respond when the attack takes place. The main objective of this research is to predict the cyber attack in the field of data engineering. Here the proposed algorithm Random Forest (RF) gave best accuracy of 95.154%, precision of 92%, recall of 94%, f1 score of 93% and AUC-ROC of 0.97 which is high compared to other existing algorithms.

Keywords: Cybersecurity · Machine Learning · Anomaly Detection · Threat Detection · Predictive Analysis · RFE · DBSCAN · Random Forest · Isolation Forest · Gradient Boosting

1 Introduction

Today the nature of cyber threats is rapidly changing and posed a significant challenge for modern businesses [1] Legacy cybersecurity products often lag the ever-evolving methods with which malevolent individuals break into systems. This exposes the vulnerability that traditional methods are more concerned with post issue resolution than in preventing those issues [2]. Now more than ever, with the demands for proactive defensive response and early warning systems reaching an all-time high, advanced threat intelligence capabilities are vital. ATSIS uses innovative telecommunication technologies (real-time monitoring, big data analytics) combined with machine learning algorithms

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 35–49, 2025.
https://doi.org/10.1007/978-3-031-86305-9_4

to address this information deficit. These systems seek to identify various threats to their organizations before they become harmful. The core of the efficiency behind ATIS is its different data integration including [3]. Additional sources include system logs, network traffic and third-party threat intelligence feeds [4]. This system provides in-depth knowledge of the network background and helps prevent cyber threats by collecting data from different sources to detect individual anomalies. One of the main challenges is that IT structures generate massive data in an organization [5]. The unstructured and messy nature of this data means that it needs heavy massaging to be able to use given all the variety, veracity, and volume issues. Unprocessed data must be cleaned, normalized, and processed to prepare it for mining [6].

This is the most important phase, as better the input data more accurate will be our threat detection model Fast solution to the image-to-image search problem requires features extraction capability [7], where ATIS needs. This method seeks to find suitable attributes in the preprocessed data which can characterize risks. It uses some complex algorithms for getting these details. Traits such as these can be detected by external threat intelligence feeds, abnormalities in system records or odd behavior with network traffic. Inputting the traits are done in Machine learning models to execute this methodology of threat identification. The data gathered from the IoT sensors are used to identify hazards, which is processed through machine learning algorithms and da it then leads to ATIS [8].

The previous post discussed the supervised and unsupervised learning methods how these make it easier for AI model to detect whenever an intrusion attempt happened on Windows based computer. Nodes which increase detection capabilities of this system. An unsupervised learning model is one that can smartly learn new threats with anomaly detection whereas supervised learning models analyze pattern of known behavior using historical data. When these techniques all work together, then ATIS can catch and neutralize any cyber threat. As ATIS must process and analyze a large amount of real time data at any given moment, the necessity for Big Data Technologies cannot be overemphasized [14].

This is done with the assistance of technologies such as Apache Hadoop and Spark that can help in managing extensive datasets which then allows security analysts to function more effectively ensuring rapid threat detection and response. Real-time cyber security should be accompanied by the ability to parse through data in real time. Longer times to detection create opportunities for harrowing outcomes. One of the key components to these ATISs is predictive analytics functionality; this part of the system anticipates common cyber-attack methods by looking at historical data and identifying trends. This feature offers resilience to organizations capable of honing their pro-active defense.

The ATIS early warning alerts show exactly what the issue was, which systems were impacted and a potential fix. These alerts were created to ensure that cybersecurity teams could properly respond in a timely and effective manner. Thanks to ATIS, the cybersecurity industry as it exists today has arguably come a very long way. Using real-time monitoring, a knowledgeable team in Israel can detect and identify potential threats by using big data analytics alongside machine learning to create an early warning system. Thus, ATIS can prove to be invaluable for any organization in today's world where

cyber threats are growing and getting sophisticated as it supports real-time alerting which instantly raises a flag on possible undetected threat or a coming danger. Future developments would probably focus on improving the predictive capabilities of these systems and making them suitable with a wider array cyber-threat, all to foster a safer digital eco-space.

This reserch article has been divided into following secttions. Section 1 represents a detailed introduction part, Sect. 2 represents a detailed survey on latest papers, Sect. 3 represents proposed model, Sect. 4 represents results obained by proposed work compaed with existing algorithms and finally Sect. 5 represents conclusion and future work.

2 Literature Survey

Dalal et al. which also provides [10] that the new architectural paradigm of running Sandboxing, using machine learning to predict cybersecurity risks. This technology aims to make systems more secure by exposing potential dangers like malware, fraud attempts and unauthorized access. The model protects the user and enterprise systems from different categories of cyber-attacks by predicting malwares ahead with proactively training them using machine learning. To improve the ant penetration system, it can be also integrated with sandbox execution to execute & evaluate possible threats in a more isolated environment.

Sapienza et al. [11] that the proposed system analyzes cybersecurity-related behavior on social media sites such as Twitter and interactions between malicious actors on the dark web to automatically notify users of any cyber-threats, whether real or imagined. Almost 84% of the 661 alerts that the system sent from 2016 to 2017 were related to threats that were either present or imminent. The article employs case studies to elucidate the system's efficacy, logic, and process. These cases illustrate the system's superior threat detection capabilities, which are demonstrated by its capacity to anticipate distributed denial of service (DDoS) attacks and promptly identify data intrusions.

Nimavat Dhaval et al. [12] proposed a new technique called MQTT (Message Queuing Telemetry Transport) protocol which helps better in predicting Quality of Service (QoS) attack. Nimavat Dhaval et al. [13] proposed a new technique called AWS MQTT (Message Queuing Telemetry Transport) which works together in predicting Denial of Service (DoS) attacks within an IoT environment.

Gowthami, G. & Priscila, S. S [14] proposed a new concept of intrusion detection using Convolution Neural Network (CNN) which gave best results in predicting anomaly detection inside a network. DevaKirubai, J. C. & **Priscila, S**.S [15] proposed a new technique in predicting intrusion in the cyber network using Artificial Neural Network (ANN). From the results obtained its proved that ANN works better in predicting the intruder (Table 1).

Table 1. Literature Survey Observed

Authors	Result Observed
Dalal et al., [10]	Cyber-attacks detected
Sapienza et al., [11]	DDoS attacks detected
Nimavat Dhaval et al. [12]	QoS attacks detected
Nimavat Dhaval et al. [12]	DoS attacks detected

3 Proposed Model

3.1 Data Collection and Integration

As a threat analysis tool, the backbone of ATSIS is data acquisition and integration-enabling it to pull in various data sources for conducting an exhaustive examination. Initially, data is systematically gathered from numerous sources like internal network traffic, system logs as well as external threat intelligence feeds. This, in turn increases the system's ability to identify prospective threats because each characteristic has a unique perspective of what is happening on the security landscape. Most importantly, it controls communication between the many systems and endpoints inside its enterprise network.

The system can pick up subtle irregularities in this data-patterns that might be the only signal alerting an organization they are under attack. Also, network traffic data is gathered by monitoring the various sensors and their positions in your networking architecture. These systems could be used to find possible irregular or harmful behaviors by collecting and analyzing data on a packet level. System logs are detailed records of events that have taken place on a system, and often contain further invaluable context. In every infrastructure the log management system is one of its vital components as it collects and stores data from servers, firewalls intrusion detection systems (IDS), etc. This log data can be analyzed to identify potential security breaches that disrupt normal operations.

A single log management system allows for consolidation and analysis of logs from multiple sources in a more possible way. Other threat intelligence sources supplement the internal data by giving a worldwide view of vulnerabilities and threats (public & private) known to exist. Many streams are available for IOCs, malware signatures attack patterns and IAT. Additionally, this will help the system in keeping up to date with latest threat trends as well it can improve its detection capabilities by combining external insights about threats with internal data. All these data sources require a strong intake mechanism that can be tapped into.

This takes all the format of data using which it can be inspected, analyzing directly from where certain type grabbles to require by shaping connectors, application program interface (API) and Extracts is used for changing data. ETL means Extracting data from a variety of sources, transform to ensure compatibility and restructuring the contents such that it meets requirements then Load in a central repository. This central repository manages the storage and retrieval of large data sets with big-data technologies like Hadoop or Spark. The result of ATIS's systematic data collection and integration (our

pipelines that collect external threat intelligence feeds, system logs, network traffic etc.) is our Cyber Threat Intelligence Dataset or CTID. Close examinations of this data during subsequent analysis and threat detection procedures will help organizations better identify (and anticipate) cyber threats, but also paint a more complete picture of the security environment as it exists today. Proposed System Architecture Fig. 1.

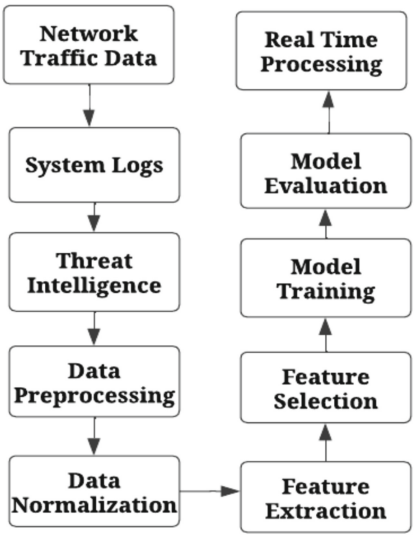


Fig. 1. Overall Proposed Architecture

3.2 Data Preprocessing and Normalization

Data Cleansing and Preprocessing: Data normalization is done as a precursor which ensures that data collected is then further ready for analysis by removing noisy elements from it and so increasing accuracy (here the term cleansing urges to remove noise). The next step is cleaning and normalizing the raw data, which means after all those different streams of data have been aggregated into what we will call a Cyber Threat Intelligence Dataset (CTID) it can be cleaned up to make sure the reliability, useability. The initial phase of preprocessing is Data Cleansing in which errors, inconsistencies or inaccurate information are identified and corrected from the dataset. This operation consists in to retroactively duplicate records, correct erroneous inputs and fill the missing values. Redundancy is Supposed to happen when multiple streams are recording the same thing.

By eliminating this duplication, we are now able to show volunteers for each event one by one. To validate the analysis results To correct any incorrect data entries, such as incomplete log entries or damaged data packets. They are also quite common in raw big data collection and can be dealt with through means like imputation or deletion based on the context it is used and how important is that missing value. Normalization is the next step after cleaning in which data should be transformed into a format that it must meet

the standards, following this comes normalization. This stage is crucial as the data are structured in a variety of different formats, and collected from many sources. You might think normalization as the standardization of data types, formats, and measure units.

This is needed to ensure accuracy and worth of Temporal Analysis by followed annotate timestamp format over entire data. If we go back to IP addresses, they are also standardized in hundreds of ways. The second important element of normalization is featuring scaling. In this operation, the numerical features are scaled into a common range using methods such as min-max scaling or z-score normalization. Scaling helps to minimize the effect on too big features in machine learning models and analysis. In fact, one common approach is to scale the volume of network traffic (which can vary dramatically) down so that it may be more easily processed and analyzed by machine learning algorithms. Data enrichment In data preprocessing is the process of generation or inclusion of additional information to provide more context for analysis.

Possible ways include enriching the logs with geo-location data based on IP addresses or consolidating log entries to threat intelligence indicators. Through Enrichment to make the dataset be more valuable when analyzing and detecting threats Modern Extract, Transform and Load (ETL) frameworks provide complete pipeline orchestration with preprocessing and normalization operations. The sum of these technologies is instrumental in contributing to assure the integrity and quality of CTID by securing that different operations as data cleansing, normalization or enrichment work effective and reliable. ATIS ensures that the data in the CTID is accurate, consistent and complete by searching for dirty characters to clean, standardize and enrich all of this dataset. A Feature-rich dataset is very important for extracting features efficiently, training ML models and automated detection of threats in real-time. It also makes the system more capable of providing early warnings about potential security threats.

3.3 Feature Selection

The Advanced Threat Intelligence System (ATIS) takes the Cyber Threat Intelligence Dataset (CTID), which contains a huge number of cyber threat intelligence features and transforms it to actionable insights using two key steps: feature extraction, then selection. These are required to detect and predict the threats accurately, which include identifying; data processing by extracting most significant attributes from processed data (cleaning). The feature extraction starts by choosing the appropriate CTID attributes that could indicate a cyber threat. This is all based on raw data which encapsulates outliers in activity system records, network behaviour trends and threats originated from the outside. Data mining uses complex algorithms to process data sets, and statistical methods that can be applied across millions or billions of records.

This could be anything from a high volume of network traffic preventable logon activity, to multiple access attempts where they should not have otherwise occurred. Often, these patterns will be used to signify malicious behavior, such as unauthorized access attempts or data exfiltration, and even Distributed Denial of Service (DDoS) attacks. Together these processes form the extraction of Anomalies, behavior analytics and Time series analysis. Time-series analysis essentially is recognizing related observations over time, where repeatable types of operational state changes are occurring inconsistently

through time. Researchers used anomaly detection tools to find atypical cases. In addition to feature extraction, behavioral analytics flags outliers in user and system activity that may reflect account takeovers or insider attacks. Feature selection is useful after the acquisition of features to increase threat detection effectiveness.

That is a fundamental step to improve the ml model performance, reduce computational complexity and dimensionality. With the usage of different feature selection methodologies like Mutual Information, Recursive Feature Elimination (RFE), Chi-square testing this is used to assess how much valuable each feature contributes. For this, the methods assess both attributes with their respective connection and contribution to general prediction accuracy score as well to find which ones provide significant light into potential threats. Recursive Feature Elimination removes weakest features from the feature set continuously till only the most important features are left and model starts performing better.

Chi-square tests pick variables with some strong relation to the target variable and then they determine how independent are these traits. Find the best features: Mutual information (Mutual info, a statistic tracking how much of each feature contributes to empower understanding about an output variable) After this step, the selected attributes must be transformed and then standardized for their input into machine learning models. This ensures that the models can accurately parse input data and output predictions. By filtering through the most important variables for detecting subtle patterns and anomalies that herald new cyber threats, these models increase the system's early warning capabilities. By leveraging the CTID as a database, ATIS can identify and predict threat events with high precision rate due to employing an elaborate process of feature extraction, selection. As a result of this, the most crucial components are stressed by it and thereby develops an operational security mechanism that enables best in class cyber threat identification and mitigation.

3.4 Feature Selection

After cleansing and classifying the Cyber Threat Intelligence Dataset (CTID), machine learning models are then utilized by another half - The Advanced Threat Intelligence System (ATIS) for producing actionable threat predictions. This procedure has an objective of building, training and testing machine learning models that can predict or detect cyber threats accurately. To begin with we need to choose the right machine learning algorithms specifically used on cyber threat-related data. Hazards are sorted with unsafe training data using supervised learning techniques. In this work we use Random Forest as algorithm. The models are trained using CTID historical data, which includes labeled examples of benign and dangerous acts. This grants algorithms with the deep learning ability to identify patterns and signs connected to many kinds of cyber threats, giving it the power classify unstructured data correctly. Researchers employ unsupervised learning techniques, such as anomaly detection models and clustering algorithms, to identify unforeseen hazards. Some algorithms that may be employed to assess data that has not been previously labeled include Autoencoders, DBSCAN (Density-Based Spatial Clustering of Applications with Noise), and K-Means. These algorithms are designed to identify anomalous patterns and events. These models are particularly adept at identifying zero-day attacks and other intricate threats that lack any known signatures.

During the training phase, the processed and normalized CTID is provided to the selected machine learning models. In this scenario, the models are continuously modifying their parameters to reduce the predicted margin of error.

Pseudocode

1. Load and preprocess the dataset (CTID)
   1.1 Clean the data (handle missing values, remove duplicates)
   1.2 Normalize the data (scale features to a standard range)
   1.3 Split the data into training and testing sets (80% training, 20% testing)
2. Feature extraction and selection
   2.1 Identify relevant features (e.g., network traffic volume, login attempts)
   2.2 Select features based on importance (using techniques like Recursive Feature Elimination)
3. Train the Random Forest model
   3.1 Initialize the Random Forest classifier with parameters (e.g., number of trees)
   3.2 Train the model using the training set
      For each tree in the forest:
      3.2.1 Draw a random bootstrap sample from the training set
      3.2.2 Build a decision tree using the bootstrap sample
      3.2.3 At each node, select the best split among a random subset of features
      3.2.4 Grow the tree to its full depth (or until a stopping criterion is met)
   3.3 Aggregate the predictions from all trees to make the final prediction (majority voting)
4. Evaluate the model using the testing set
   4.1 Predict the threat labels for the testing set
   4.2 Calculate evaluation metrics (accuracy, precision, recall, F1-score, AUC-ROC)
5. Deploy the trained model for real-time threat detection
   5.1 Integrate the model with real-time data processing pipelines
   5.2 Continuously monitor network traffic and system logs
   5.3 Predict and alert for potential threats in real-time
6. Update and retrain the model periodically
   6.1 Collect new data and integrate it into the dataset
   6.2 Retrain the model to adapt to new threat patterns

Model performance is improved and overfitting is prevented through the implementation of techniques like cross-validation and hyperparameter tuning. Dropping the dataset into multiple phases is as an approach to shield models from overfitting new information; these methods are all part of cross-validation. The most accurate predictions may require the modification of certain model parameters, such as learning rate in Gradient Boosting or number of trees in a Random Forest. The performance of a model is judged through validation after training. These measures include recall, accuracy, precision, F1-score and Area under the Receiver operating characteristic curve (AUC-ROC), effectively categorizing threats by labelling them as either confirmed or not. The process of validation assures the models maintain a low level of false negatives standing as well as possible to prevent creating too many overfit/false positive problems. Once validated, the models are

live within our ATIS framework to both detect and predict hazards in real-time. Once integrated with the system's data ingestion and processing pipelines, models may be configured to evaluate incoming data streams from the CTID continuously after deployment. Different Models enable cybersecurity teams to respond more quickly and faster if there is any potential threat detected (Proactive Response). Machine learning models are constantly retrained by feeding them with more data. A major part in their ability to adjust models over time with an evolving threat environment and thus for identifying novel threats. ATIS utilizes highly sophisticated artificial intelligence driven machine learning models to greatly enhance proactive cybersecurity approach, and provide alerts in real time. Below is the pseudocode for the proposed work.

3.5 Real-Time Big Data Processing

The real-time big data processing capabilities of ATIS significantly improve its ability to analyze large volumes of data from the Cyber Threat Intelligence Dataset (CTID) and provide timely threat alerts. To achieve this capability, it is essential to integrate streaming data frameworks with modern big data technologies that facilitate continuous data intake, processing, and analysis. The process of data collection commences with the acquisition of pertinent information from a diverse array of sources, such as external threat intelligence feeds, system records, and network traffic. Using technology Apache Kafka these data streams are integrated in real-time using robust data pipelines. Kafka, for example, captures all critical data elements and makes them accessible for rapid processing due to its capacity to handle high-throughput, low-latency data input.

The data is received by a real-time processing engine, such as Apache Spark Streaming or Apache Flink, after it has been ingested. In order to facilitate the concurrent and continuous implementation of intricate analytical tasks, these engines are engineered to process data in micro-batches or stream-processing modes. For example, Spark Streaming has the capability to analyze large datasets in real time to identify trends, anomalies, patterns, or other indicators of potential cyber threats by partitioning them into smaller, more manageable segments. The data transformation, enrichment and parsing all these operations are affected by the processing engine to a large extent. It allows our raw data to be transformed or "parsed" into a more structured form, so that it can actually TO BE analyses. Data is transformed so that it standardized and normalized to the same level as across all sources. In other words, you might use the CTID to enrich IP addresses with their geo-location content or combine log entries that have been linked by a recognized threat indicator. This processed data streams then feed into advanced analytics models and machine learning model trained on historical data.

These models look at the data for potential signs that may indicate a malicious activity in the context of real-time processing system. Anomaly detection and classification models will alert administrators to hazard signs in a network data like the one done by Example. If there are any risks, then these judgments will trigger rapid alarms. The system is highly scalable, fault-tolerant, and reactive due to a computational model employed. It uses distributed computing clusters that can scale horizontally by adding more nodes for increasing data needs. A more versatile storage system, like Amazon Simple Storage Service (S3) or Hadoop Distributed File System (HDFS), is used to efficiently store and access a large amount of data. Real-time dashboards and alerts in

the system give cybersecurity teams the ability to respond to attacks immediately. SIEMs typically integrate with automated alerting systems to notify users using channels like SMS or email. Dashboards to see data presented with focus on important metrics and warnings ATIS helps in providing risk management by collecting and acting on big data (identifying hazards to address them in real-time).

3.6 Predictive Analytics and Threat Alerting

To predict and provide timely warnings of potential cyber attacks, the Advanced Threat Intelligence System (ATIS) employs Cyber Threat Intelligence Dataset (CTID). Using advanced analytical models enables the system to examine both historical and real-time data, identify patterns based on this data over time, predict future outcomes according these patterns as well potential indicators of a possible threat that are setting off alarms in security team heads. Predictive analytics relies on machine learning algorithms that have been trained on the Cyber Threat Information Database (CTID). It keeps a record of all cyber incidents, network behaviors and threat indicators concerning the planet. The data is analyzed using methods to identify patterns in behavior and time, such as regression analysis or time-series forecasting, along with machine learning classifiers. For example, consider the case of time-series forecasting models. They can predict network traffic and system behavior trends using historical data to save them. Any abnormality that these models detect could propose the presence of a new hazard. A typical example is that regression analysis could identify which categories of network activity have a probability for being correlated to the likelihood of a cyberattack. Machine learning classifiers that have been trained to identify patterns associated with specific hazards using both historical and real-time data may be able to identify early warning signs of aggressive behavior.

By utilizing the CTID's retrieved features, classifiers such as neural networks, decision trees, and random forests can accurately predict the probability of a variety of threat scenarios. Alerts will be promptly transmitted after the identification of potential hazards by the predictive analytics of the system. These alerts are highly actionable due to their provision of precise information regarding the peril, its consequences, and methods for mitigating it. The probability of the risk being genuine is determined by the threat type, affected systems, confidence ratings, and time markers. By integrating the alerting mechanism with real-time monitoring systems, warnings are transmitted immediately upon the identification of a potential threat. Email, SMS, and connections through SIEM systems to transfer the alarms are the measures that may be used to make the technology systems capable of sending the notification. Since the alarms may be automated in the platform, the desired method is dependent on the recipient's preferences and need for assurance that the notification reaches them; moreover, the multi-channel approach ensures that security staff is notified promptly, regardless of jurisdictions. Finally, the security analytics platform includes a dashboard with real-time information on the development of threats. Dashboards display crucial data such as IP address, username, event category, among other variables in a manner that presents the trends clearly and aids speedy judgment and reaction.

4 Results and Discussion

4.1 Data Set Information

Background: The Cyber Threat Information Dataset (CTID) is a for extensive set of data, consisting different kinds of sources like. - A list if external threat information feeds, - Network traffic datasets and - System event records. On collecting this data, it amounts to a few terabytes and the entire process runs at around 750 GB per day. The data capture should extend to IOCs from reputable threat intelligence sources, system event logs, user-authentication logs and network traffic. The dataset has been cleaned and harmonized to ensure uniformity across different types of data, sources. By training and validating employee machine learning models on this large dataset, ATSIS can detect a wide variety of cyber threats. One of the Table 2 against components dataset is shown here [16].

Table 2. Dataset Information

Data Source	Data Type	Volume (GB/day)	Collection Method
Network Traffic	Packet Data	500	Network Monitoring Tools
System Logs	Event Logs	200	Centralized Log System
Threat Intelligence	Indicators of Compromise (IoCs)	50	External Threat Feeds

4.2 Output Metrics Evaluation

The effect of the ATIS models is evaluated by standard metrics with common usage, such as AUC-ROC, F1-score, recall, accuracy, and precision. The Random Forest classifier yielded dominant values: F1-Score of 93%, Recall Score - 94%; Precision Score-92% and Accuracy-95%. For anomaly detection, the Isolation Forest model has an accuracy of 93%, a precision of 91%, recall in average with sensibility at class [1] -positive- is $< 90\%$ and F1-score $=$ $\sim 90\%$. They show models' ability to detect/change the threat categories. The real-time processing engine built using Apache Spark Streaming processed huge volumes of data in a throughput at the rate of 200 MB/s with an average latency threshold within 100 ms. These results suggest that ATIS utilizes CTID to detect and predict cyber threats in a timely manner with high precision. Their high accuracy and recall rates demonstrate the reliability of models for recognizing true hazards with little false positives. Security staff can act proactively on potential issues because alarms and threat detection happen almost immediately (if data is analyzed in real-time). By using predictive analytics, which enhance the system's ability to detect and prevent

future attacks as well this gives us a broader cyber security that represents how far we are from thinking in terms of the full cyber ecosystem. Overall, the results of this case are impressive: The combination of real-time big data processing and modern machine learning allows improving cyber protections in a company. Table 3 and Fig. 2 shows each algorithm's metrics value.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

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

$$F1 - Score = 2 * \left(\frac{Precision * Recall}{Precision + Recall} \right) \tag{3}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Table 3. Output Metrics

Model	Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Threat Classification	Random Forest [Proposed method]	95.154	92	94	93	0.97
Anomaly Detection	Isolation Forest [8]	93	91	90	90	0.95
Trend Prediction	Gradient Boosting [10]	94	93	92	92	0.96

Equation (1) shows the precision formula, Eq. (2) depicts the recall formula, Eq. (3) depicts the F1-Score formula and Eq. (4) depicts the accuracy formula, where TP means True Positive, TN means True Negatives, FP means False Positives, FN means False Negatives. Figure 3 depicts the confusion matrix.

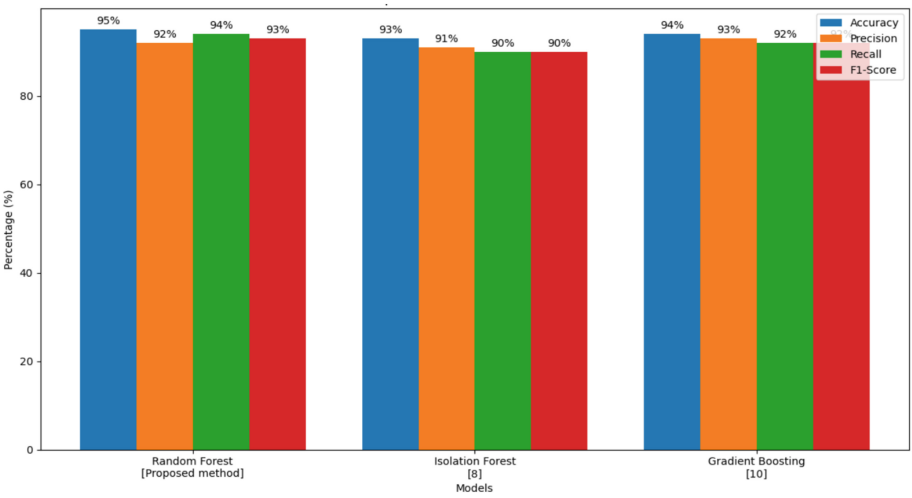


Fig. 2. Comparison metrics graph

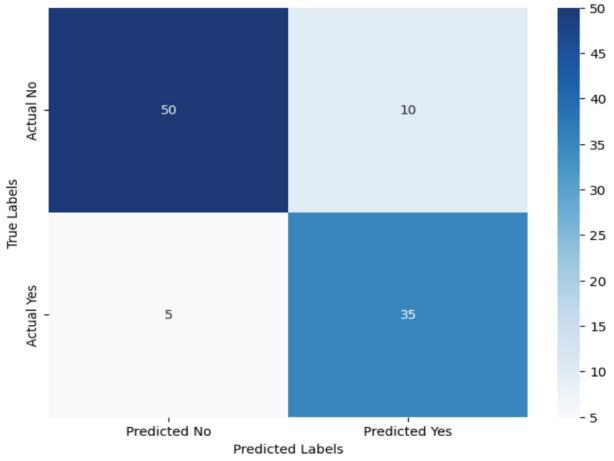


Fig. 3. Confusion Matrix

5 Conclusion and Future Work

The sophisticated Threat Intelligence System (ATIS) establishes a strong foundation for the early detection and prevention of cyber assaults by integrating predictive analytics, sophisticated machine learning models, and real-time big data processing. ATIS is capable of preprocessing and standardizing data because of the Cyber Threat Intelligence Dataset (CTID). Robust algorithms are employed to extract pertinent properties in the detection and forecasting of cyber threats. A real-time alerting system enhances a company's cybersecurity posture by ensuring prompt responses. It is possible to implement

proactive threat management and simultaneously protect critical information systems from cyber-attacks through this comprehensive approach.

References

1. Nunes, E., et al.: Darknet and deepnet mining for proactive cybersecurity threat intelligence. In: IEEE Conference on Intelligence and Security Informatics ISI, pp. 7–12 (2016)
2. Robertson, J., et al.: Darkweb Cyber Threat Intelligence Mining (2017)
3. Shakarian, J., Gunn, A.T., Shakarian, P.: Exploring malicious hacker forums. In: Cyber Deception Building the Scientific Foundation, Springer, pp. 261–284 (2016)
4. Rele, M., Patil, D., Boujoudar, Y.: Integrating artificial ıntelligence and blockchain technology for enhanced US homeland security. In: 2023 3rd Intelligent Cybersecurity Conference (ICSC), San Antonio, TX, USA, pp. 133–140 (2023)
5. Moll, J., Yigitbasioglu, O.: The role of internet-related technologies in shaping the work of accountants: new directions for accounting research. Br. Acc. Rev. **51**(6), 100833 (2019)
6. Sanclemente, G.L.: Reliability: understanding cognitive human bias in artificial intelligence for national security and intelligence analysis. Secur. J. **35**(4), 1328–1348 (2022)
7. Ahmed, S., Hossain, M.F., Kaiser, M.S., Noor, M.B.T., Mahmud, M., Chakraborty, C.: Artificial intelligence and machine learning for ensuring security in smart cities. In: Data-Driven Mining Learning and Analytics for Secured Smart Cities: Trends and Advances, pp. 23–47. Springer, Cham (2021)
8. Waseem, M., Adnan Khan, M., Goudarzi, A., Fahad, S., Sajjad, I.A., Siano, P.: Incorporation of blockchain technology for different smart grid applications: architecture prospects and challenges. Energies 16(2), 820 (2023)
9. Da Xu, L., Lu, Y, Li, L.: Embedding blockchain technology into IoT for security: a survey. IEEE Internet of Things J. **8**(13), 10452–10473 (2021)
10. Dalal, K.R., Rele, M.: Cyber security: threat detection model based on machine learning algorithm. In: 2018 3rd International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, pp. 239–243 (2018)
11. Sapienza, A., Bessi, A., Damodaran, S., Shakarian, P., Lerman, K., Ferrara, E.: Early warnings of cyber threats in online discussions. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), New Orleans, LA, USA, pp. 667–674 (2017)
12. Nimavat Dhaval, M., Raiyani Ashwin, G.: A study on MQTT protocol architecture and security aspects within IoT paradigm. In: Balas, V.E., Semwal, V.B., Khandare, A. (eds.) Intelligent Computing and Networking. Lecture Notes in Networks and Systems, vol. 632. Springer, Singapore (2023). https://doi.org/10.1007/978-981-99-0071-8_6
13. Nimavat Dhaval, M., Raiyani Ashwin, G.: Detecting and preventing DoS attacks within IoT environment using AWS IoT Core". In: Castillo, O., Sudhakar Babu, T., Aluvalu, R. (eds.) Pervasive Knowledge and Collective Intelligence on Web and Social Media. PerSOM 2023. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 517. Springer, Cham (2024). https://doi.org/10.1007/978-3-031-66044-3_5
14. Gowthami, G., Priscila, S.S.: Convolution neural network-based efficient development of ıntrusion detection using various deep learning approaches. In: Paramasivan, P., Rajest, S., Chinnusamy, K., Regin, R., John Joseph, F. (eds.) Explainable AI Applications for Human Behavior Analysis, pp. 224–241. IGI Global (2024). https://doi.org/10.4018/979-8-3693-1355-8.ch014

15. DevaKirubai, J.C., Priscila, S.: Artificial neural network-based efficient cyber hacking detection system using deep learning approaches. In: Paramasivan, P., Rajest, S., Chinnusamy, K., Regin, R., John Joseph, F. (eds.) Explainable AI Applications for Human Behavior Analysis, pp. 205–223. IGI Global (2024). https://doi.org/10.4018/979-8-3693-1355-8.ch013
16. https://www.kaggle.com/datasets/zunxhisamniea/cyber-threat-data-for-new-malware-attacks

Reconciliation - Backdoor Access Finding Strategies with Legacy Applications

Bhushan B. Chavan[1(✉)], Harsh Soni[2], Lakhichand Khushal Patil[3], and Kalpesh A. Popat[4]

[1] IAM, MGM Resorts International, 6770 S Edmond St, Las Vegas, NV 89118, USA
bchavan@mgmresorts.com
[2] 991 US Highway 22 West, Bridgewater, NJ 08807, USA
[3] Fergusson College (Autonomous), Pune, Maharashtra, India
lakhichand.patil@fergusson.edu
[4] Marwadi University, Rajkot, Gujarat, India

Abstract. In the intricate realm of Identity and Access Management (IAM), practitioners frequently grapple with a host of prevalent challenges that impede the process of access reconciliation. A prime hurdle is the intricate task of pinpointing inconsistencies and discrepancies within user access rights spread across a plethora of systems and applications. The inherently complex and ever-evolving nature of modern IT landscapes, where access levels are meticulously assigned based on varied roles and responsibilities, exacerbates this challenge. Compounding this issue is the lack of a centralized overview of user access rights, which obstructs the timely detection and resolution of conflicts, such as unwarranted or excessive access privileges. These conflicts not only heighten security risks but also amplify the organization's vulnerability to unauthorized access. Addressing these issues necessitates the adoption of robust mechanisms aimed at identifying and rectifying access discrepancies, including regular access reviews and automated access certification processes. In this paper,

We wrote access reconciliation algorithm to identify backdoor, Manual Error, Timing error exceptions. It is found that the outcomes are satisfactory compared to manual reconciliation and as a result these techniques can be integrated in designing an expert tool and helpful in Audit's, Technical hardenings.

Keywords: Access provisioning · Access de-provisioning · Manual Provisioning · Access credentials · Authorization · Computer security · cybersecurity · information security

1 Introduction

Identity and Access Management significance has grown as the digital information security domain has become more dominant than ever, safeguarding data securely and enabling smooth user operations on various platforms. Necessary to manage user identities and control access to organization resources, IAM systems encounter several problems as network size and intricacy expand. These include scalability restrictions, integration barriers, concerns about privacy, and stringent compliance prerequisites. As a

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 50–66, 2025.
https://doi.org/10.1007/978-3-031-86305-9_5

result, the success of tough security mechanisms and operational effectiveness relies on confronting these. The paper's objective is to investigate technological details regarding lot of IAM access reconciliation, unrevealing diverse procedural implementations contributing to strengthening IAM systems, and examine real-world case studies and best practices revealing the excellence of original strategies. Paper help to implement automation of Access reconciliation tool in Banking & Finance, Hospitality, Media, Gaming Industries. In Major enterprises legacy tools are used across different line of business. With these legacy tools due to incorrectly handling SOP or errors in Manual operation (due to Human read mistakes) create Audit issues. This paper simply gather data in three data sets, 1. Applications User Access List, 2. Legacy Request Tool Data, 3. Human Resource Data i.e. PI data and perform access reconciliation. Backdoor exception is really critical in terms of SOX, PCI applications since corporate employee can do any un-authorized activity with backdoor access. With this access reconciliation paper & tool we have benefits to pass the Audit and prevents Frauds. The Paper / Tool on access control help enterprise's operational efficiency requirements through automation. The analysis will leverage several stakeholders' perspectives to provide an elaborate picture of the cogent challenges and trailblazing solutions within the IAM sector. Solution can be considered for both Corporate and customer identity [1].

2 Technological Challenges in IAM Access Reconciliation

2.1 Scalability Concerns

As organizations burgeon and diversify, scalability emerges as a formidable challenge in the realm of IAM access reconciliation. The task of managing access rights for an ever-increasing roster of users, systems, and applications escalates in complexity and demands substantial resources. The traditional manual methods of access reconciliation, often cumbersome and error-prone, falter under the pressures of scalability. To effectively address these scalability issues, organizations are compelled to embrace automated IAM solutions that can adeptly manage large-scale access reconciliation tasks. These solutions, harnessing the power of identity analytics, machine learning, and artificial intelligence, automate the identification and resolution of access discrepancies. Through the automation of these processes, organizations can significantly enhance operational efficiency, curtail the risk of errors, and ensure access reviews and certifications are conducted promptly [2].

Current situation cannot last any longer. Large organizations, which seem to have not faced the problem of error-prone access reconciliation, will shortly feel some symptoms of the problems disturbing them. So to solve the situation, IAM access reconciliation processes should be automated. Solutions strong in identity analytics, machine learning and AI technologies should be implemented for managing big data and for automating the correction of all detected inconsistencies. As a result, these processes will be managed more efficiently and much faster with difference implemented, access reviews and certifications will be carried out in time, and the risk of human error will be minimized.

2.2 Integration with Existing Systems

The second would-be technological obstacle in effective access reconciliation is the integration of IAM systems with pre-existing systems and applications. Many business organizations operate within a complex information technologies landscape that presupposes a mixture of legacy systems, cloud- based applications, and third-party platforms, all of which require access management as part of IT administration. To ensure the effective reconciliation of access, it is essential to remove the barriers to integration and ensure superior interoperability of IAM systems with a variety of environments [3].

The challenges of integration are associated with issues like data synchronization, user provisioning or deprovisioning, and the management of access controls across systems that differ from each other. To avoid the challenges, a suitable integration framework and standards should be created by organizations. The frameworks help to ensure that interaction of systems that are a part of IAM and any other platforms. You keep talking mostly about access control system, but not integration systems, and more general topics like this is harmonized so that the data and access information are exchanged smoothly [4].

2.3 Data Privacy and Security Risks

IAM access reconciliation, at its core, raises fundamental issues about data privacy and security. IAM systems, large repositories of sensitive information and private access data, must be secure under any and all circumstances. Anything less could well lead to unregulated access, data leaks and misuse, as well as additional legal complications and sanctions [5].

To protect the data and rights of access for the user, organizations need to put in place a suite of security measures, such as data encryption in all conditions, secure access controls, and regular assessment of the security of systems. Also, compliance with existing data protection laws and standard industry practices should be ensured to guarantee the confidentiality and security of user information [6].

2.4 Compliance and Regulatory Constraints

One of the challenges of IAM access reconciliation is compliance with regulatory and industry standards requirements. If healthcare, financial, government and other sectors that have to comply with strict compliance requirements come into consideration, compliance with various compliance frameworks that govern data access control, protection, and privacy will be a challenge.

The challenge will be addressed by introducing explicit access control regimes, periodically performing access audits, and documenting the access policies to comply with sector-specific or international standards. Moreover, since legislation and policies are not static, enterprises should reconcile access policies periodically. Finally, violating compliance may cause severe penalties and reputation loss for the company participating in the sector [7].

3 Methodological Approaches to Access Reconciliation

3.1 Automation in IAM

The incorporation of automation within the realm of Identity and Access Management (IAM) significantly refines the process of access reconciliation [8]. Through the adoption of proposed automated tools, organizations not only diminish the necessity for manual intervention but also curtail the likelihood of errors and bolster operational efficiency. Such systems are adept at identifying and rectifying any mismatches between user access rights and entitlements automatically, ensuring adherence to organizational policies and compliance mandates. Moreover, automation facilitates the swift revocation of access for individuals who either depart the organization or transition to different roles, thereby mitigating the risk of unauthorized data exposure [9]. Implementing RBAC control will help for assigning access based on distinct roles predefined according to job functions and organizational structure. However, application which don't support automation of access provisioning and de-provisioning, required security administration team to do manual provisioning and de-provisioning of users access. Example in USA different states has gaming board requirement on access provisioning and de-provisioning and don't allow automations unless they approved by gaming boards. All major applications provisioning and de-provisioning go through Manually process and as per Gaming Commission rule Security administration team has to compliant with Gaming license. In USA all major states has gaming commissions like Nevada Gaming Commission etc. There are high chances to errors in Manual provisioning and Audit findings. With this automation we will able to resolve these GAPs and helps auditing and monitoring techniques plays a vital role in the maintenance of access integrity within IAM [10].

Security Objectives - defines three types of security objectives for information and information systems.

- Confidentiality
- Integrity
- Availability

Impact on Enterprise and Users

Security Categorization information system = {(confidentiality, impact), (integrity, impact), (availability, impact)}, where the acceptable values for potential impact are LOW, MODERATE, or HIGH (Refer Table 1).

4 IGA Tool and Sequence Diagrams

Below is diagram which shows the sequence of provisioning and de-provisioning access with different use cases.

Type of Access Control is Detective Access Control

A detective control attempts to discover or detect unwanted or unauthorized activity. Detective controls operate after the fact and can discover the activity only after it has

Table 1. Various Authorization Level

Authorization Level			
Impact Categories	Priority 1	Priority 2	Priority 3
Inconvenience, distress or damage to standing or reputation	LOW	MOD	HIGH
Financial loss or agency liability	LOW	MOD	HIGH
Harm to agency programs or public interests	N/A	Low/Mod	HIGH
Unauthorized release of sensitive information	N/A	Low/Mod	HIGH
Personal Safety	N/A	Low	MOD/HIGH
Civil or criminal violations	N/A	Low/Mod	High

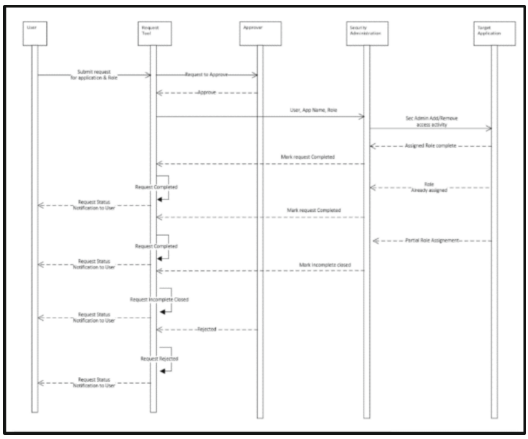


Fig. 1. Sequence Diagram Which Shows How Request Process Flow Works

occurred (Refer Figs. 1, 2, 3 and 4). Example reviews of access, backdoor access, Manual errors, Timing errors [11].

Access Operations

1. Every enterprise activity in IAM includes below
2. Provisioning access to new employees
3. Adjusting access rights to employees being transferred
4. Assisting employees with access such as a forget password, accounts locked, Disable account
5. Removing employees access from separation with enterprise

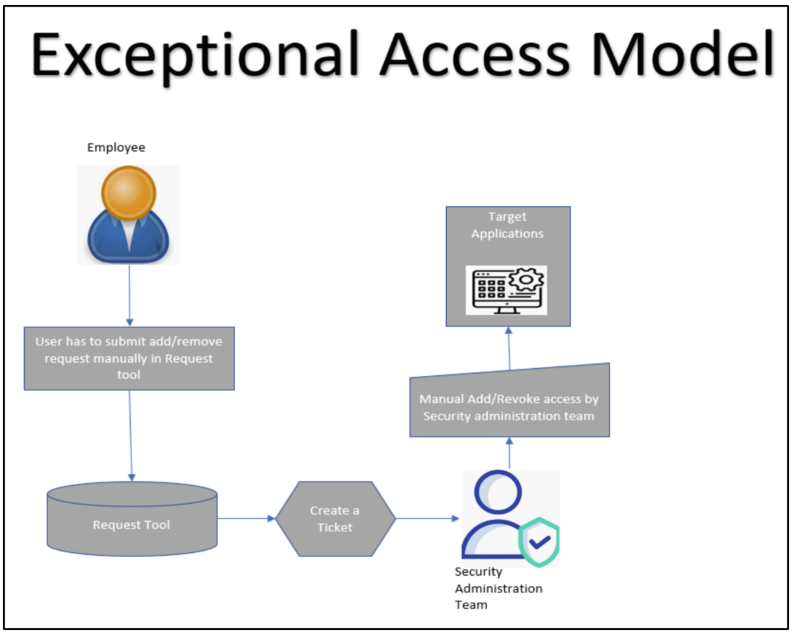


Fig. 2. User Requesting Access Exceptionally

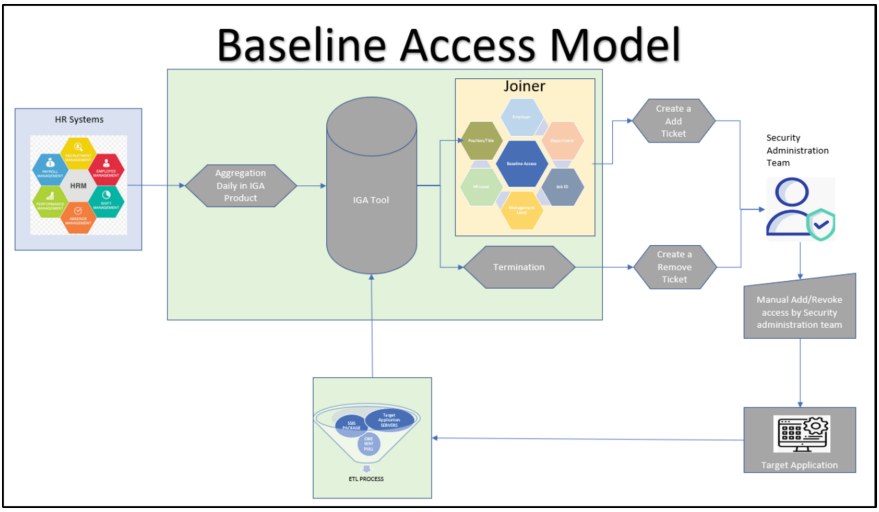


Fig. 3. User Access with Baseline Rule

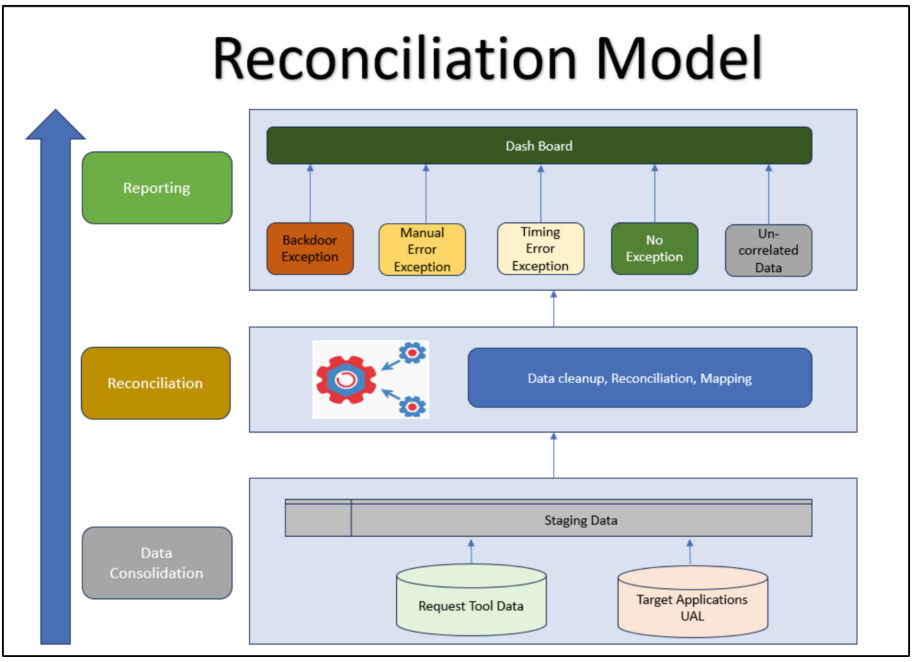


Fig. 4. Access Reconciliation Model

5 Algorithm

Access Reconciliation Theory.

Access Requests:
An access request is made for a user to grant or revoke access for a role. Each access request has a timestamp associated with it.

$$ar = (u, r, op, t) \mid u \in U, r \in R, op \in \{"grant", "revoke"\}, t \in T$$

where-

1. U is the set of users.
2. R is the set of roles.
3. op is the operation type.
4. T is the set of all possible timestamps between the time when app was first configured until today, i.e.,

$$T = [t_{onboarded}, t_{current}]$$

Roles and Entitlements:
A role is associated with an application and can contain multiple entitlements

$$r = (app_r, e_r) | e_r \subseteq E, app_r \in A$$

where E is the set of entitlements and A is the set of applications.

Every entitlement has a name (identifying a security group within an application context) and is associated with a provisioning engine.

$$e = (name_e, p_e) | p_e \in P$$

where P is the set of provisioning engines.

A and P aren't mutually exclusive i.e., some applications may be their own provisioning engines.

$$\exists X : (X \subseteq A) and (X \subseteq P)$$

Users' Access List:
An access may be granted to a user through an association with an entitlement. This association (called a grant event) will also have a timestamp.

$$ge = (e, u, t) | u \in U\ , e \in E, t \in T foreveryge \in UAL$$

where UAL called 'Users Access List' is the set of all grant events i.e., $ge \in UAL$

When an access is revoked, the corresponding grant event is removed from the set.

$$UAL \leftarrow UAL \backslash \{ge\}$$

Delta UAL and Delta Requests:
If AR represents the set of all access requests i.e., $ar \in AR$

Then we select a set of recent requests from a given time that represents last date of reconciliation, up until current time.

$$\Delta AR = \{(u, r, op, t) \in AR | t_{last} \leq t < t_{now}\}$$

Similarly, we select a set of recent grant events

$$\Delta UAL = \{(u, r, t) \in UAL | t_{last} \leq t < t_{now}\}$$

Final Resolution:
We expand the set ΔAR for each entitlement in each role of the access requests. This gives us a set where we have one entry for each entitlement requested as part of a role.

$$\Delta AR_e = \bigcup_{ar \in \Delta AR} \{(u, e, op, t) | e \in e_r \in r \in ar\}$$

The function final resolution is then applied to this set to get the latest access request for each unique combination of user and entitlement. The idea here is to consider the "grant" or "revoke" operation whichever was the most recent.

$$f(x) = \{(u, e, op, t) \in x | \forall (u', e') \in x, \quad (u, e) = (u', e') \Rightarrow t > t' \vee op = op'\}$$

For the sake of simplicity, let's call this Delta final resolution, ΔFR

$$\Delta FR \leftarrow f(\Delta AR_e)$$

Reconciliation Goal:
We're looking for results for each user and entitlement across both ΔUAL and ΔFR

$$ReconciliationResults = \{Reconciliation(u, e) | (u, e) \in \left(f(\Delta AR_e) \bigcup \Delta UAL\right)\}$$

$$Reconciliation(u, e) = \begin{cases} UnauthorizedAccess \\ Pending\text{Re}vocation \\ PendingGrant \\ AuthorizedAccess \end{cases}$$

1. Unauthorized Access: There exists a combination of user and entitlement in users' access list for which there is no matching user and entitlement in access requests with a "grant" operation.

$$\exists(u,e) \subset ge \in \Delta UAL : \neg\exists(u,e) \subset ar \in \Delta FR : ar_{op} = "grant"$$

2. Pending Revocation: There exists a combination of user and entitlement in users' access list for which there is a matching user and entitlement in access requests with a "revoke" request.

$$\exists(u,e) \subset ge \in \Delta UAL : \exists(u,e) \subset ar \in \Delta FR : ar_{op} = "revoke"$$

3. Pending Grant: There exists a combination of user and entitlement in access requests with a "grant" request for which there is no matching user and entitlement in users' access list.

$$\exists(u,e) \subset ar \in \Delta FR : ar_{op} = "grant" : \neg\exists(u,e) \subset ge \in \Delta UAL$$

4. Authorized Access: There exists a combination of user and entitlement in access requests with a "grant" request for which there is a matching user and entitlement in users' access list.

$$\exists(u,e) \subset ar \in \Delta FR : ar_{op} = "grant" : \exists(u,e) \subset ge \in \Delta UAL$$

Reconciliation Algorithm: Objects in Algorithm

1. Potential cases of unauthorized access (ω): We make use of this set when we make a strong assumption that all new grant events result in unauthorized access.

2. Confirmed cases of unauthorized access (Ω): We use this to denote all of our reconciliation results which are confirmed and continue to be unauthorized access.
3. Potential cases of pending revocation (ρ): We create these cases for all new requests with "revoke" operation so that they can be checked again later for persisting access.
4. Confirmed cases of pending revocation (R): We use this to confirm a scenario for a request with "revoke" operation that has persisting access.
5. Confirmed cases of pending grant (G): We use this to confirm a scenario for a request with "grant" operation that hasn't yet been fulfilled.
6. Confirmed cases of authorized access(Φ): We use this to denote a perfect reconciliation result.
7. Reconciliation result (rr): A record that stores the result of Reconciliation(u, e) as shown above

6 Algorithm Steps

1. Declare empty sets:

$$\omega = \emptyset, \rho = \emptyset$$

2. Iterate Over Each Grant Event in ΔUAL:

If there exists a confirmed case of pending grant, remove that case from the memory as the access is now granted. If there does not exist such a case, create a new potential case of un-authorized access.

$$\forall ge \in \Delta UAL,$$

$$if\ \exists rr \in\ G\colon (u,e)_{ge} = (u,e)_{rr}\ then$$
$$G \coloneqq G\ \setminus \{rr\}$$
$$Else\colon$$
$$\omega \coloneqq \omega\ \cup \{rr_{ge}\}$$

3. Iterate Over Access Requests in Δ FR:
   a. For all access requests with "grant" operation, see whether access was granted or not. If it was granted, create a case of authorized access.
   b. If it was granted as part of the new grant events, delete the corresponding potential case of unauthorized access.
   c. If the access was not granted, create a new confirmed case of pending grant.
   d. For all access requests with "revoke" operation, check in entire users' access list whether or not access still persists. If it does, create a confirmed case of authorized access. If not, create a
   e. potential case of pending revocation.
4. Mark all potential cases of unauthorized access and pending revocations as confirmed:

$\forall ar \in \Delta FR,$

$If\ ar_{op} =$ "grant" then
{ $If\ \exists ge \in \Delta UAL : (u,e)_{ar} = (u,e)_{ge}\ then$
{ $\Phi := \Phi \cup \{rr_{ar}\}$
$If\ \exists rr \in \omega: (u,e)_{rr} = (u,e)_{rr_{ar}}\ then$
$\omega := \omega \setminus \{rr\}$
}
$Else$:
$G := G \cup \{rr_{ar}\}$
}

$If\ ar_{op} =$ "revoke" $then$
{ $If\ \neg\exists ge \in UAL: (u,e)_{ar} = (u,e)_{ge}\ then$
$\Phi := \Phi \cup \{rr_{ar}\}$
$Else$:
$\rho := \rho \cup \{rr_{ar}\}$
}

Since we have accounted for all requests, any undeleted potential cases are confirmed cases of unauthorized access

$$\Omega := \Omega \cup \omega$$

Since we check the entire users access list, we will mark all cases to be confirmed. Having a separate reference allows us to skip over these records in our final step of the algorithm.

$$R := R \cup \rho$$

5. Final Iteration Over Unauthorized and Pending Revocations:

For all previous cases of unauthorized access and pending revocations, check whether entry was deleted from UAL or not. If it was, remove that case from its respective set.

$\forall rr \in (\Omega \setminus \omega) \cup (R \setminus \rho),$
$if\ \exists ge \in \Delta UAL : (u,e)_{rr} = (u,e)_{ge}\ then$
{ $If\ rr \in \Omega\ \ then\ \ \Omega := \Omega \setminus$ rr

$Else\ If\ rr \in R\ then\ R := R \setminus rr$

}

7 Algorithm Analysis

Fixed Assumptions

1. **Ordering**:
   - ΔFR and Both UAL and ΔUAL are native to the IGA software, and therefore sorted sets that can perform search based on user and entitlement with at least the performance of a balance binary tree.
   - Ω, R, G, ω and ρ can be implemented as sorted sets too but will rely, in most implementations, on the indexing of the storage medium such as a relational DB. In the interest of determining a worst case analysis, we will assume these sets to be unsorted and therefore the search to be a brute-force linear traversal.
2. **Sizes**:
   - UAL size: N
   - ΔUAL size: n (a subset of UAL)
   - ΔFR size: m
   - Larger sets like Ω, R and G (confirmed historical cases): Q
   - Smaller sets like ω and ρ (potential cases that are fresh variables in the algorithm): q

Complexity Analysis

1. Iterating Over ΔUAL
   a. Complexity: $O(nQ)$, due to searching in the larger unordered set G (size Q).
2. Iterating Over ΔFR
   a. For "grant" operations:
      i. Binary search in ΔUAL (size n): $O(\log n)$ per operation
      ii. Linear search in smaller set (size q): $O(q)$ per operation.
      iii. Total for each "grant" operation: $O(\log n + q)$
   b. For "revoke" operations:
      i. Binary search in UAL (size N): $O(\log N)$ per operation.
      ii. Total complexity for "revoke": $O(m \log N)$
3. Final Iteration Over Larger Sets Ω, R, and G:
   a. Binary search in UAL (size N): $O(\log N)$ per entry
   b. Considering Q as the size of these larger sets, total complexity: $O(QlogN)$

Overall Complexity
Combining the complexities of the major steps, the overall complexity of the algorithm is:

$$O(nQ) + O(m(\log n + q)) + O(m \log N) + O(Q \log N)$$

Since $q \leq n$ complexity can be rewritten as:

$$O(nQ) + O(m(\log n)) + O(m \log N) + O(Q \log N)$$

Since $n < N$, complexity can be rewritten as:

$$O(nQ) + O(m \log N) + O(Q \log N)$$
$$= O(nQ) + O(\max(m, Q) \log N)$$

This comprehensive analysis considers the interactions between the various set sizes and the efficiency of the search operations within these sets. The complexity is influenced by the sizes of N, Q, M and n, and the types of searches performed (binary in UAL and ΔUAL, linear in the unordered sets).

Observations

This comprehensive analysis considers the interactions between the various set sizes and the efficiency of the search operations within these sets. The complexity is influenced by the sizes of N, Q, M and n, and the types of searches performed (binary in UAL and ΔUAL, linear in the unordered sets).

1. With smaller m and n quantities, i.e., more frequent access reconciliation that generates smaller ΔUAL and ΔFR, the first term $O(nQ)$ and second term $O(m \log N)$ gets reduced, thereby resulting in a shorter process.
2. With a smaller set of confirmed cases, i.e., Q, the first term $O(nQ)$ and the third term $O(q \log N)$ is reduced, thereby improving performance.
3. If binary search can be performed on sets of confirmed cases (size Q) the first term would also become linearthmic i.e., $O(n \log Q)$ thereby improving performance.

Summary

Access reconciliation involves comparing a set of grant events with a transformed set of access requests to determine the status of user access. The process includes a series of transformations to align the access requests set for effective comparison. The reconciliation algorithm iterates over these sets and previous results to validate and update access statuses. Its efficiency hinges on the sizes of the input sets, the entire user access list, and the number of confirmed historical cases in various categories of access reconciliation results. With optimized data storage and retrieval strategies, the algorithm can potentially achieve near-linearithmic time complexity. However, an efficient implementation must also account for space complexity, balancing speed with resource utilization [12].

8 Case Studies and Best Practices

8.1 Case Study Analysis

Reviewing case studies offers critical information on the anticipated outcomes of IAM access reconciliation strategies in various real-life contexts. For example, one needs to consider a case associated with multinational corporations. Namely, the example demonstrates how IAM access reconciliation has been implemented in the context of a large multinational corporation. The case is related to a variety of business units and their multiple locations. With the help of this approach, the company's goal was to obtain consistent access management to meet regulatory requirements. The organization has

used a set of automated tools and implemented role-based access control. As a result, the major achievement includes a significant decrease in the time and costs associated with providing/restricting access [13].

Moreover, the case study emphasizes the importance of regular monitoring and auditing of the system. The company conducted audits to detect and eliminate the existing intervals in the access to sensitive data, eliminate inactive accounts, and take other measures to eliminate other potential ranges to the facility. For these purposes, monitoring and auditing were vital to enhance access policies and improve the overall effectiveness of the IAM system.

8.2 Success Stories in Various Industries

IAM access reconciliation strategies have been successfully applied across all industries, from healthcare and finance to manufacturing. In the healthcare sector, hospitals and other organizations related to the industry have applied IAM protocols to secure patient records and other sensitive information. The strategy ensures that only personnel authorized to do so can reach these data at any level of security.

Banks and other institutions in the finance industry have also embraced IAM access reconciliation to increase security and comply with regulatory requirements. They have implemented role-based access management and automated the process of managing access rights. As a result, banks have decreased the number of potential risks related to safety breaches more effectively and deliver more streamlined and efficient operations. Companies operating in the manufacturing industry have also used IAM access reconciliation to protect their production systems and secure intellectual property. They implemented very restrictive access to their systems and data, often requiring an authorization procedure to collect data from the factory floor. Regular reviews of these measures have allowed these businesses to maintain the confidentiality of their projects and prevent potential breaches in security [14].

8.3 Lessons Learned

The analysis of different case studies and real-world applications has led to the overview of several crucial lessons. Firstly, a profound understanding of an organization's business processes and data access requirements appears to be the most essential condition for an effective access reconciliation process. Secondly, automation provides significant gains, including a reduction in the number of workflow events combined with a lower probability of the occurrences of human-made errors. Thirdly, it is crucial to establish an incessant process of monitoring and auditing for early detection of discrepancies in access and security threats. Finally, based on the experience of numerous organizations, RBAC policies appear to be worth using to determine users' access to each specific "resource based on their role and responsibility within the organization". In sum, industrial experience shows that access reconciliation is a complex process that must be conducted by the stringent efforts of IT and business stakeholders [15].

8.4 Benchmarking and Metrics

To ensure ongoing improvement, it is crucial to work out effectiveness measuring of IAM access reconciliation initiatives. This could be done not only by comparing organizations to industry standards to set their KPIs but also in terms of regular assessment. Such down-to-the-ground understanding of the effectiveness allows companies to identify the areas that need further improvement [16].

8.5 Future Directions and Innovations

The IAM access reconciliation domain is developing rapidly due to the advance in technology and constant changes in the security landscape. Some of the promising trends and developments of the future include:

Key metrics: time for access provisioning, and deprovisioning, the duration of access review cycles, number of access discrepancies, and compliance violations

Integration of blockchain – this technology will change IAM forever by creating a centralized, shared, append-only record of user identities and access rights.

Advanced machine learning and AI that will not only automate reconciliation processes but analyze them and identify any suspicious activity and also predict potential security risks.

Further developments in data privacy and security, in line with the new very strict regulation.

More focus on user experience – all the tools should be easy and simple to use.

Wider utilization of cloud-based IAM since they are more scalable, flexible, and cost-efficient.

With the help of these approaches, organizations will be able to remain competitive and adjust their IAM to the specific digital environment of the employees. Overall, through the analysis of the case studies, and best practices, one may gain the insights of which strategies and methods were efficient and what are the existing challenges and issues in the domain of IAM access reconciliation. The cases show that automation, RBAC model, continuous monitoring, and policy improvement work well and the emerging technologies such as AI and blockchain, have the potential to change the existing approaches. Metrics and benchmarking, in the meantime, should be used when developing the strategies and evaluating them. Future trends, also, suggest the opportunities for progress in the future. Overall, the review of previous implementations and the adoption of emerging practices may help organizations to develop their IAM properly [9].

9 Comparative Analysis

(Due to privacy/confidential information we can't disclosed application name's and enterprise study come from. This analysis for first 1000 Users from each application).

10 Conclusion

10.1 a. Technological Challenges and Methodological Approaches

The study & implementation has acquired a better understanding of the technological challenges faced by IAM systems in the current legacy dynamic technological environment.

10.2 b. Methodological Approaches

The results of the methodological approaches investigated during the study are associated with crucial strategies to overcome the challenges regardless of the discussion above. The results include automation, implementation of role-based access control RBAC, effective audit and monitoring, definition and review management policies, and artificial intelligence, machine learning and improving the performance of the access reconciliation process.

10.3 c. Case Studies and Best Practices

Although the information acquired through the review of case studies in several organizations, as well as the implementation of controls applied across different industries, has not been connected to specific results, I have managed to uncover effective implementation examples and identify significant mistakes. Thus, the idea related to the different exceptions was developed as a part of the review to measure the effectiveness of access reconciliation.

10.4 d. Importance of IAM Acknowledgement

Overall, the paper allowed me to understand the significance of IAM access reconciliation for enhancing the level of data security and the quality of user interaction. The methodological processes support the findings and results of the exceptions for the paper.

Author's Contribution Statement. All the authors have contributed equally to the design and implementation of the research, to the analysis of the results, and to the writing of the manuscript.

Ethical and Informed Consent for Data Used. As authors, we give full permission for the use of the research study's data and ethical considerations.

Data Availability and Access. Since the research study in question comprised confidential information, it was stored in a secure location and was only accessible upon request. We guarantee that the research study's use of data won't reveal any personal information.

Conflict of Interest. The authors have no conflicts of interest to declare. All co-authors have seen and agree with the contents of the manuscript and there is no financial interest to report. We certify that the submission is original work and is not under review at any other publication.

References

1. Giaretta, A., Pepe, S., Dragoni, N.: UniquID: A Quest to Reconcile Identity Access Management and the IoT, pp. 237–251 (2019)
2. La Rue, C.J., et al.: Groups 4 retirement: a new intervention that supports well-being in the lead-up to retirement by targeting social identity management. J. Occup. Organ. Psychol. **97**(1), 1–26 (2024). https://doi.org/10.1111/joop.12458
3. Parent, M.C., Wille, L.: Heterosexual self-presentation, identity management, and sexual functioning among men who have sex with men. Arch. Sex. Behav. **50**(7), 3155–3162 (2021). https://doi.org/10.1007/s10508-021-01968-z
4. Mikula, T., Jacobsen, R.H.: Identity and access management with blockchain in electronic healthcare records. In: 2018 21st Euromicro Conference on Digital System Design (DSD), pp. 699–706 (2018). https://doi.org/10.1109/DSD.2018.00008
5. Patil, L.K., Popat, K.A.: Comparative Analysis of Several Approaches of Encoding Audio Files, vol. 1760 CCIS. Springer Nature Switzerland (2022)
6. Hardjono, T.: Federated authorization over access to personal data for decentralized identity management. IEEE Commun. Stand. Mag. **3**(4), 32–38 (2019). https://doi.org/10.1109/MCOMSTD.001.1900019
7. Sharma, D.H., Dhote, C.A., Potey, M.M.: Identity and access management as security-as-a-service from clouds. Procedia Comput. Sci. **79**, 170–174 (2016). https://doi.org/10.1016/j.procs.2016.03.117
8. Patil, L.K., Popat, K.A.: Design and Implementation of Multilayer Encryption for Audio File Security, pp. 179–191 (2024)
9. Indu, I., Anand, P.M.R., Bhaskar, V.: Identity and access management in cloud environment: mechanisms and challenges. Eng. Sci. Technol. an Int. J. **21**(4), 574–588 (2018). https://doi.org/10.1016/j.jestch.2018.05.010
10. Caron, J., Markusen, J.R.: CISM_Book (2016)
11. Stewart, J.M., Chapple, M., Gibson, D.: CISSP (ISC)2 Certified Informat - James M. Stewart Sybex Wiley Edition 7. Wiley
12. Sharma, A., Sharma, S., Dave, M.: Identity and access management- a comprehensive study. In: 2015 International Conference on Green Computing and Internet of Things (ICGCIoT), pp. 1481–1485 (2015). https://doi.org/10.1109/ICGCIoT.2015.7380701
13. Ghaffari, F., Gilani, K., Bertin, E., Crespi, N.: Identity and access management using distributed ledger technology: a survey. Int. J. Netw. Manag. **32**(2) (2022). https://doi.org/10.1002/nem.2180
14. Olabanji, S.O., Olaniyi, O.O., Adigwe, C.S., Okunleye, O.J., Oladoyinbo, T.O.: AI for identity and access management (IAM) in the cloud: exploring the potential of artificial intelligence to improve user authentication, authorization, and access control within cloud-based systems. SSRN Electron. J. (2024). https://doi.org/10.2139/ssrn.4706726
15. Aboukadri, S., Ouaddah, A., Mezrioui, A.: Machine learning in identity and access management systems: survey and deep dive. Comput. Secur. **139**, 103729 (2024). https://doi.org/10.1016/j.cose.2024.103729
16. Fan, X., Chai, Q., Xu, L., Guo, D.: DIAM-IoT: a decentralized identity and access management framework for internet of things. In: Proceedings of the 2nd ACM International Symposium on Blockchain and Secure Critical Infrastructure, pp. 186–191 (2020). https://doi.org/10.1145/3384943.3409436

Enhancing Cyber Hacking Breach Classification Through Integrated Preprocessing and Feature Extraction Techniques

J. Christina Deva Kirubai[1(✉)] and S. Silvia Priscila[2]

[1] Department of Computer Applications, Bharath Institute of Higher Education and Research (BIHER), Chennai, India
christina.cs@bharathuniv.ac.in

[2] Department of Computer Science, Bharath Institute of Higher Education and Research (BIHER), Chennai, India
silviaprisila.cbcs.cs@bharathuniv.ac.in

Abstract. The increasing risk of cyber attacks necessitates the use of advanced and effective techniques for quick detection and categorization. Through the integration of preprocessing approaches and feature extraction techniques, the present study aims to improve cyber hacker breach categorization in terms of accuracy, precision, and recall. To be more precise, Z-Score standardization is used as a stage of preprocessing to normalize the data, and robust feature extraction is achieved by Independent Component Analysis (ICA) after that. For thorough breach categorization, the retrieved characteristics are subsequently combined using Multilayer Perceptron (MLP), Decision Trees (DT), and Naive Bayes classifiers. The research task assumes that Z-Score normalization, ICA, and MLP working together will produce better outcomes than any of them working alone. Experiments conducted on a wide range of cyber-hacking situations reveal that the suggested method works well, with improved memory, accuracy, and precision. The results highlight the value of integrating both preprocessing and feature extraction methods and highlight the opportunities for real-world security-related applications enabling active breach prevention and detection. From the outcomes attained, the proposed MLP yields accuracy rate of 91.50%, Precision of 0.90, Recall of 0.89. The tool used is Jupyter Notebook and the language used is python.

Keywords: Cybersecurity · Hacking breaches · Classification · Multilayer Perceptron · Decision Tree · Naive Bayes · Preprocessing · Z-Score · Feature extraction · ICA (Independent Component Analysis)

1 Introduction

Using integrated preprocessing and feature extraction approaches to increase threat detection systems' accuracy and efficiency is one way to improve the classification of cyber hacking breaches. Preprocessing methods like data cleaning and normalization standardize input and remove noise from raw data before preparing it for analysis.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 67–79, 2025.
https://doi.org/10.1007/978-3-031-86305-9_6

Meanwhile, feature extraction methods find and isolate the most pertinent indicators of possible breaches. Combining these strategies allows businesses to produce datasets that are more informative and well-refined, which improves classification algorithm performance. This integrated approach guarantees that the classification system can adjust to changing attack patterns and developing vulnerabilities in addition to aiding in the proper identification and categorization of cyber threats.

At the moment, there are many other terms used to refer to cybercrime, including computer crime, e-crime, Internet crime, etc. The risks associated with cybercrime are rising daily. In addition to a shortage of documentation at the relevant offices for a variety of causes, including victims' presumptions regarding police reaction, the lack of knowledge of IT act on cybercrimes, and victims' incapacity of recognizing whether they are being attacked there is at present no foolproof, systematic, dependable instrument on analyzes of cybercrimes. Preventing and categorizing cybercrimes may also be aided by raising public awareness of the issue, identifying trends in electronic crime, and maintaining area-specific cybercrime statistics [1].

Regulatory compliance, incident response, and effective risk management all depend on the classification of cyber hacking breaches. Organizations can use it to priorities hazards, deploy resources effectively, and put customized reaction plans into action. Adequate categorization guarantees adherence to reporting mandates and facilitates effect evaluation, assisting in conveying the gravity of violations to relevant parties. It also helps with strategic planning by helping to spot patterns and guide preventative security measures. Classification also affects legal actions, insurance claims, and communications both internally and externally, which improves an organization's overall cyber security posture and resilience [2].

Cybercrime is becoming more and more prevalent in the internet era. Since there is certainly not enough data to solve a cybercrime case in earlier, an algorithm for ML is essential. With this framework, information can be precisely categorized through analysis, and by considering attributes into account, groups may be anticipated. The definitive aim is to increase the network's safety performance to protect it from intruders [3]. In order to identify, track, protect, and identify IoT devices, applications, users, and data about online threats, a number of enabling techniques using AI and ML have recently been developed [4].

2 Literature Survey

According to a 2020 cybersecurity ventures study, the annual cost of cybercrime is expected to approach $6 trillion by 2021. The body of research on cybercrime crimes is currently lacking in computer approaches for cybercrime prediction, particularly for data that is not organized. As a result, RupaCh et al., 2020) suggest a versatile computational framework that makes use of approaches to assess the rate of computer crimes at the organizational level within a nation and aid in the classification of cybercrimes. By combining statistical techniques with security analysis, crimes using India-based integrated data are analyzed and categorized, which can be either organized or unorganized. Assessment analysis states, that identifying the infractions precisely with the highest possible accuracy, is the model's key strength [1].

One of the most significant tools to defend Internet of Things, or IoT, gadgets from intrusions is the detection of cyber security vulnerabilities in these gadgets as they may be used. However, this is becoming a more difficult task. Through the application of algorithms based on ML on a variety of datasets, namely IoT23, Sarah Bin Hulayyil, et al., 2023 perform a thorough study to explore the approaches as well as technologies utilized in determining vulnerabilities in IoT contexts. Here the ML pipeline for identifying IoT vulnerabilities is detailed, and the most frequent potential drawbacks of IoT systems are examined on each tier [5].

AnkitaKumari et al.'s 2021 study focuses on methods for utilizing different categorization algorithms to forecast hazards to network security. The four main categories of techniques for classification which are covered here include DT, NB, KNearest Neighbors, and logarithmic regression. It assesses the effectiveness of the previously described classification techniques to determine the threats [6].

Data safeguarding objectives are achieved through the employment of numerous techniques, nevertheless, breach detection still takes a lot of work. Mohammed NaifAlatawi et al., 2022, have developed PSO-GA (PSO-based GA), a hybrid methodology that combines swarming thinking with evolution approaches for feature selection, using information in a structured manner. The greatest precision of 100% was attained in the present study using brute force, SQL injection, and PortScan, demonstrating the usefulness of the suggested approach in cyber security programs [7].

Frank Cremer et al. in 2022 conducted research on cyber security and cyber risk mitigation from academia and business and paid special attention to the accessibility of data. After using the methodical process, 5219 cyber-peer-reviewed publications were identified from the original search, providing 79 distinct data. To better understand, quantify, and make out cyber risks, cybersecurity experts and insurance businesses will benefit from the assessment and classification of the resultant data [8].

The creation and implementation of a methodology for finding webpages that breach data is the subject of M R Prathap et al.'s 2021 publication. The RF technique was selected for this particular endeavor because it is the most effective technique for classifying an enormous quantity of letters. The undertaking consists of three primary phases: gathering data, extracting features, and executing training methods. The method of extracting features from URLs from websites is a crucial stage in this procedure. For research, a collection of 17 characteristics is retrieved. Two phases come after feature the extraction process: training and testing. A developed ML procedure will be produced at the conclusion. Future projects can take advantage of this developed algorithm. These days, a lot of websites that compromise user data are developed. This ML methodology will efficiently detect those internet addresses that are violating data [9].

3 Proposed Model

Loading a dataset into the structure is an essential step in training and assessing a cyber hacking breaches. Selecting a excellent dataset is critical and it is choose from kaggle.com. The dataset consists of 25 attributes and 40,000 records [10]. This study uses the Kaggle Cyber Hacking dataset. The complete system design is depicted in Fig. 1.

The diagram outlines a ML workflow for classifying cyber-security attacks using a dataset from Kaggle. The process begins with input data from the Kaggle cyber-security-attacks dataset. This data undergoes preprocessing, specifically using Z Score normalization, to standardize the data. Following preprocessing, feature extraction is performed using ICA to reduce dimensionality and isolate the most informative features. These extracted features are then fed into various classification algorithms, including MLP, DT, and Naive Bayes, to train models capable of predicting cyber-attacks. The final step is output prediction, where the performance of these models is evaluated using scales like precision value, recall rate, and accuracy to determine the effectiveness of the classification.

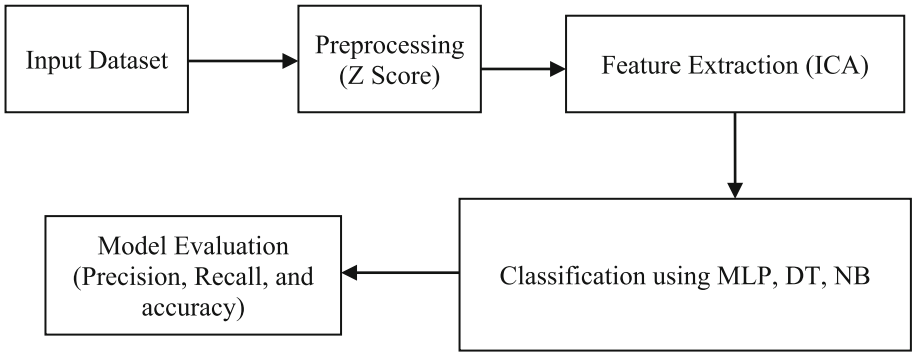


Fig. 1. Overall Recommended Framework

3.1 Preprocessing Using Z Score

There are various reasons to preprocess data, but one that is particularly important in a cyber-hacking breach classification as the latter involves working with cybersecurity data which can be both complex and very different. Proper preprocessing can significantly improve the accuracy and consistency of a set classification models used in cyber threat detection and response. Preprocessing has a major role in preserving the quality and integrity of data for analytics especially in cybersecurity domain. Since these datasets have noise, incomplete information and many relevant irrelevant details require much preprocessing in the form of feature selection/engineering or using techniques to deal with noise like Outlier treatment etc. These methods which we are going to discuss will cleanse the data and make more accurate, which eventually be a reliable resource for our model development. High-quality data sets enhance the precision of true security incidents detection increases and can also lower chances of a false alarm which gives rise to negative or positive outputs from the model [11].

In this classification method for cyber hacking, as we previously said the workflow is to normalize the dataset features by Z-score normalization process. This will help to normalize the data making sure that it has 0 mean and a standard deviation of one so it improves model training by de-focusing outliers.

Z-scores are sometimes also called “standard scores Z-score normalization is one of the simplest ways to normalize data, consisting in dividing a dataset by its standard deviation. By this the relationship between one data point and another is unchanged but are moved to a scale that can be compared. It is especially useful when working with features having different scales or units.

Here is the formula for how to calculate Z score of a data point x:

$$Z = \frac{x - \mu}{\sigma} \tag{1}$$

where:

Z denotes the data point Z-score value.

x represents the data point’s initial value.

M denotes the dataset mean value.

Σ indicates the dataset.

Cybersecurity data is characterized by features of different scales and units, or time-based metrics. By standardizing these attributes, we scale all of them to follow a normal distribution (mean = 0 and SD = 1). The reason of standardization is that it provides a scale to the algorithm so any feature does not disturb learning because its value can be large. Standardization removes the unfair advantage that some features may have over others and it helps the model to learn better, which in turn generates more accurate predictions.

In cybersecurity, datasets can be high dimensional and encompass numerous potential features that could help detect breaches. Methods like ICA, a preprocessing technique are used to decrease this dimensionality by selecting out the most informative features. Taking root of the quadratic difference squared it and hence compacts to simpler modelling resisting falling into over-fitting. Since it only concentrates relevant features, the model is going to better generalize and be more accurate in detecting cyber threats.

Preprocessing helps in increasing accuracy and reliability of classification models which are deployed on cyber security industry. Preprocessing: Preprocessing is used to clean standardize, and optimize the data set, so that models are trained with high-quality based on this data. These metrics help other performance testing scales, including areas such as precision, recall and accuracy - the fundamentals of whether cybersecurity measures are effective. The improved detection rate helps organizations to pinpoint threats more significantly and faster, therefore reducing an ideal behavior risk. Identifying the outliers, how a score fits into a distribution and statistical decisions are the main reason for using Z Score.

3.2 Feature Extraction Using ICA

It is a monumental step in the ML process, especially with respect to cyber hacking breach classification. This works to enhance the accuracy and effectiveness of models that are created for naming, detecting as well as dissecting cyber threats. It contains plenty of data, often complex and has multiple features e.g., Cybersecurity data The feature extraction methods such as ICA, can reduce this complexity by extracting and separating the important parts of raw data. We simplify and make models by focusing on features that facilitate the necessary subsequent analysis (and classification).

The pre-processed data is classified using ICA to extract vital features out of it after normalization. ICA could uncover the essential underlying patterns which are keystone in cyberattacks classification by getting mixed signals broken down into parts.

ICA is a statistical method that divides a multivariate signal into independent and additive components. Data analysis and signal processing both frequently use it.

Let X be a set of observed signals that are combinations of signals S, which are independent sources of information. This relationship can be shown as:

$$S = A \times S \tag{2}$$

where:

The observed signals are represented by the m × n matrix X, where n is the quantity of observations and m is the amount of observed signals (variables).

The independent source signals are represented by a k × n matrix, S, where k is the number of independent sources.

The linear combining of the independent sources to create the observed signals is described by the m × k mixing matrix A.

ICA aims to estimate the base signals S and the mixing matrix A from the perceived signals X alone. By isolating the most relevant features, feature extraction enhances the model's capability to identify patterns and anomalies associated with cyber breaches. This results in better accuracy of classifying the types of attacks. By having fewer but more meaningful features, the model can learn better which reduces overfitting and generalizes well for new unseen records. High-dimensional data is an expensive operation and collections of list objects. Feature extraction is used to make the computer do less work in terms of how many features per image it must process. That causes the model to be trained faster and processed more efficiently; timely responses are important in use cases such as real-time threat detection.

When the number of features is large, models are hard to interpret and understand. The feature extraction reduces the model by concentrating on only significant activity features, such that its explanation and decision-making characteristics are interpretable. This is especially pertinent in cybersecurity, where the why something was flagged can be just as important as what it was pointing to. Shadow over the Real data are crude ones with a lot of noise which will cover or remove some very important patterns in them. The process of feature extraction is used to get rid of this noise by keeping only that features which play a significant role in the prediction. This makes the model more robust i.e. less sensitive to unrelatable variations in data and focused on what really matters i.e. signals that show possible breaches. ICA helps enormously in dimensionality reduction in feature selection thus increasing the accuracy during classification.

3.3 Classification

The classification algorithms which can classify varieties of types in cyber threat by certain behavior and properties as identified from the data based is very useful for Identified Hacking Detection. To be able to detect and response a security breach on time, these algorithms help differentiate what activity is acceptable (benign) from malicious. They may prove more adept at recognizing and reigning in threats quicker, reducing

manual toil and making security systems overall function better by automating how classification takes place. Well, to maintain strong and proactive cybersecurity defenses having significant impact over the accuracy as well dependability of threat detection; proper selection of a classification algorithm is quintessential.

After extracting the features, they use three classification algorithms MLP, DT and NB to classify cases of cyber hacking.

3.3.1 MLP (Multi-Layer Perceptron)

Due to their ability to detect, too; as well as model difficult-to-read data patterns - MLP provides the essential classification framework for cyber hacking breaches. They are very good for working with high-dimensional data and they can learn complex non-linear interactions often see in cyber attack scenarios. Using historical breach data, MLPs can evolve over time to reflect changing threat patterns and leading to increased detection and categorization accuracy of attacks able be detected across the full attack lifecycle - keeping up cyber defenses.

Multi-Layer Perceptron (MLP) artificial neural networks are used in various classification problems like cyber-attack detection. An MLP consists of three or more layers, an input layer, one or additional hidden levels and also a output level layer. All the neurons in a layer are interconnected to every other neuron on subsequent layers, with each connection has an associated weight. Every neuron in the output layer and hidden layers applies an activation function to its inputs weighted sum.

During forward propagation, data enters the network then is processed layer by layer to get predictions. The output of each neuron is calculated as:

$$z = \sum_{i=1}^{n} (w_i . x_i) + b \tag{3}$$

$$a = \text{activation}(z)$$

Here:

x_i defines the i^{th} input attribute.

w_i indicates the weight connected with the i^{th} input feature.

b represents the bias term.

activation (n) denotes the activation method.

MLPs are capable of learning complex patterns in the data and make predictions by combining backpropagation, activation functions, and forward propagation. These characteristics have been exploited in the form of automatic instrumentation for cyber-attack detection and classification via iterative training that refines network parameters to minimize prediction error (i. e., increase classification accuracy).

3.3.2 DT (Decision Tree)

Given their simplicity and interpretability, after some careful hand tuning DT is still useful for classifying cyber hacking breaches. They enable an easy visualization of decision-making processes, which can provide insight into how breaches are classified according to different attributes. Since the DTs are able to work with both numerical

and nominal fields, they make it a powerful tool for analyzing variety of cyber-attack indicators. Being able to define simple decision rules helps to rapidly recognize patterns and anomalies for early detection & response. Also, DTs can be nicely integrated with other techniques to improve the quality of breach classification.

The rule used for the classification of cyberattacks using DT is based on general decision tree algorithm in this study. The entropy of a dataset provides us with this idea to measure impurity or uncertainty. The formula for calculating AUC score can only be calculated using a binary classification problem.

$$H = -p_{positive}\log_2\left(p_{positive}\right) - p_{negative}\log_2\left(p_{negative}\right) \tag{4}$$

$p_{positive}$ is the proportion of instances labeled as positive (e.g., hacking activity) in the dataset.

$p_{negative}$ is the proportion of instances labeled as negative (e.g., normal activity) in the dataset.

Information gain measures the reduction in entropy that is achieved by splitting a dataset based on an attribute. It is used in order to know which feature would split the data best. The formula for information gain is:-

$$IG = H_{parent} - \sum_{i=1}^{n}\left(\frac{N_i}{N} \times H_i\right) \tag{5}$$

where:

H_{parent} is the entropy of the parent node before the split.

n is the number of child nodes after the split.

N_i is the quantity of instances in the i^{th} child level node.

N is the entire instances in the parent level node.

H_i is the entropy of the i^{th} child node after the split.

3.3.3 NB (Naïve Bayes)

Naive bayes classifiers are widely used in cyber hacking breach classification, as it can handle large datasets with multiple features efficiently and effectively. They are effectively used with deceptive simple assumptions (the independence of features, which makes it easier to calculate and also faster the classification process). The great advantage of NB classifiers is that they can deal with both categorical and numerical data, easily scale to large datasets but also perform well on small/medium sized ones. Their probabilistic model yields well-defined results that are easy to interpret and can help provide insight on the types of breaches most likely in practice. Also, since NB classifiers can be quickly updated with new data, these are models that have high predictive adaptation which applies to our changing threat landscapes.

The NB classifier is a supervised learning type with probabilistic output which yields the possibility of some class given all input attributes. Naive Bayes employs the Bayes rule as its founding principle. This classifier uses the conditional probability of all features given a class, or p(a|b) with prior probabilities of all classes, or p(b), to calculate the posterior possibility of a class which is equal to P(b|a). The reasons for this are that

each feature make individual contributions to the overall posterior probability of a class, so naïve is employed as (Dipankar Dasgupta et al., 2020).

$$p(bla) = \frac{p(a, b)}{p(a)} = \frac{p(alb)p(b)}{p(a)} \tag{6}$$

where is the input-type vector and c = class type vector.

The combination of MLP with ICA and Z-score normalization for better classification performance. It is done by measuring the efficiency of different methods and then evaluate on parameters like recall, accuracy & precision etc. Comparative analysis is performed for choosing a best fit combination of preprocessing techniques and classification algorithms to produce an efficient cyber hacking categorization.

4 Result and Discussion

Utilizing their capability to understand patterns and anomalies identified from the data, ML approaches in this model represent a valuable trajectory towards cyber hacking categorization. In the effectiveness evaluation of this technique, standard criteria including accuracy, precisions and recall. These metrics are numerical, and show how good our model is at detecting hacking breaches without too many false positives or negatives.

The criteria used to measure the performance of a model in cyber hacking breach classification are referred as Performance metrics and it serves for getting exact values on how well or poor a detection, classification helps disgorge any security breaches. These metrics include accuracy, precision, recall and F1 score used to evaluate the model; if false positives are minimized and most breaches can be detected using these methods. These metrics inform the choice and tuning of algorithms to enhance detection performance, reduce risk, as well as improve remediation strategies. Through this a performance metrics can be analyzed to classify models that are best system along with reinforced defenses against an ever-evolving number of cyber threats.

4.1 Accuracy Analysis

In case of cyber hacking breaches, accuracy refers to how accurately a categorization model can detect instances when it comes to hacking or breach. It is the % of cases-out-of-box dataset which were correctly classified as a sum total over all instances.

The accuracy of a classification model (7) can be defined by:

$$\text{Accuracy} = \frac{\text{Number of Correct Predictions}}{\text{Total Number of Predictions}} \times 100 \tag{7}$$

Number of Correct Predictions: The quantity of cases that the model accurately classified (Table 1).

Total Number of Predictions: The entire count of occurrences inside the dataset (Fig. 2).

Table 1. Accuracy Value Vs Classifiers

Algorithms	Accuracy (%)
Naïve Bayes	86.20
Decision Tree	88.50
Multi Layer Perceptron	91.50

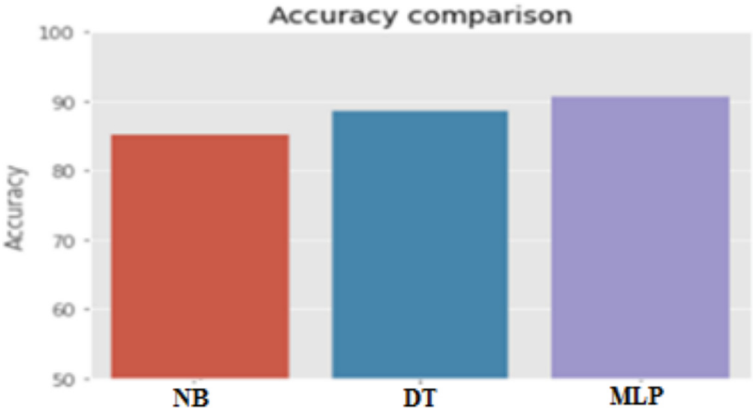


Fig. 2. Accuracy Comparison of MLP Graph with Preprocessing Techniques

The best classifier in the denoted evaluation score is MLP with a total accuracy of 91.50%. This is actually a very good accuracy which means RF correctly classifies instances overall better than NB and DT classifiers. This makes the MLP ideal for capturing intricate patterns and relationships in data which indicates that it outperforms other algorithms on this dataset.

The DT classifier follows with an 88.50% accuracy rate. It's not quite on the level of an MLP, but it still performs well. This results in a higher accuracy as compared to Naïve Bayes, which is better at correctly categorizing the data since trees can be tuned more easily with clear decision rules that are easy for humans to interpret.

The performance of NB is the worst among all, as it has an accuracy lower than Random Forests and Logistic Regression with 86.20%. Whilst it is not quite as accurate as the Decision tree and MLP classifiers, overall its accuracy seems acceptable. This smaller accuracy is indicative that NB may not be as good at capturing the characteristics of this dataset compared to other classifiers, making more errors.

4.2 Precision and Recall Analysis

Cyber-oriented Fraud detection in breaches sounds with precision and means how much good the classification model predicts. It calculates the proportion of positive cases correctly identified (that is, correct hacks) against all actual positives.

The accuracy of this classification model is formally defined by (8)

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \times 100 \tag{8}$$

Recall also known as sensitivity in the world of cyber hacking breaches, is an evaluation metric that used to measure what percentage of positive cases were caught by a classification model. It measures the percentage of all real positive occurrences—that is, projected hacker breaches—that are accurately labeled as positive instances.

The recall of a classification model can be defined using the following formula:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \times 100 \tag{9}$$

True Positives: The number of cases that were accurately identified as breaches due to hacking (Table 2).

False Negatives: The number of incidents where non-hacking actions were mistakenly categorized (predicted as non-hacking activities but are hacking breaches) (Fig. 3).

Table 2. Precision and Recall Values Vs Classifiers

Algorithms	Precision	Recall
Naïve Bayes	0.85	0.83
Decision Tree	0.88	0.86
Multi Layer Perceptron	0.90	0.89

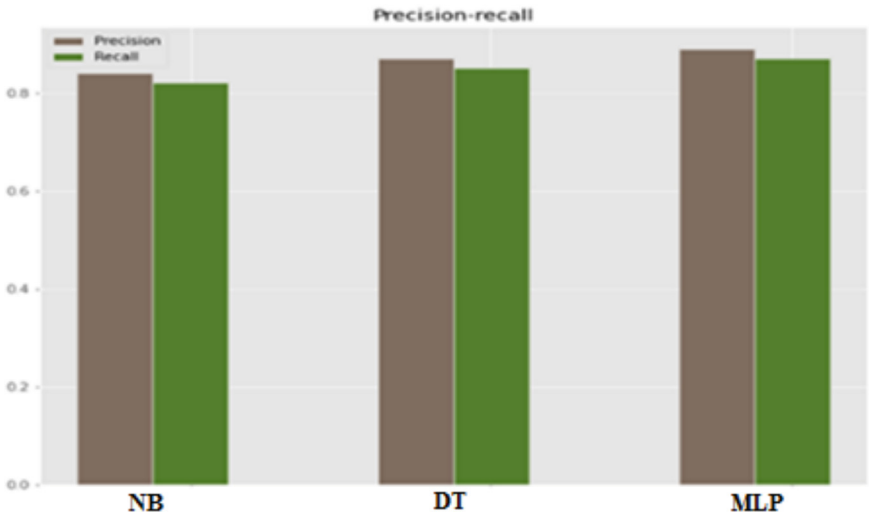


Fig. 3. Precision and Recall Comparison of MLP Model with Preprocessing Techniques

The NB classifier demonstrates solid performance with a precision of 0.85 and recall of 0.83. This means that when NB predicts a breach, it is correct 85% of the time, reflecting its accuracy in positive predictions. Its recall of 0.83 indicates that it detects 83% of all actual breaches, showcasing its ability to identify a substantial portion of relevant cases. However, compared to the other algorithms, Naïve Bayes has lower precision and recall, suggesting it may miss some breaches or include a higher number of false positives.

The results with a precision of 0.88 and recall of 0.86 are improved in the DT classifier. That means that when it predicts a breach, its right 88% of the time which is an improvement on Naïve Bayes—so less false positives. Moreover, its recall of 0.86 indicates that this model recognizes correctly approximately 86% cases considered truly true when compared to NB making classifier more efficient in detecting positive instances relevant for a breach case than the simplest approach. The DT achieves high coverage while maintaining good accuracy, outperforming Naïve Bayes models in both metrics.

The MLP classifier performs the best and has a precision of 0.90, recall at 0.89. This implies that very few false positives are produced from the MLP, guaranteeing a low error rate in predicting breaches compared to other two algorithms. Even its recall of 0.89 shows that it can detect 89% of real breaches which is one the best performance on identifying relevant cases. As such MLP results to have the highest performance in both precision and recall making it very well working model for cyber security hacking breach detection as well having high accuracy.

5 Conclusion and Future Work

This proposed approach is used to classify cyber-attacks through multiple steps, hoping that the accuracy would be enhanced by conducting preprocessing and applying ML models. We normalize the data-set by applying Z-score normalization to standardize features and keeping a consistent scale of variables. The feature was extracted from the preprocessed data using ICA and normalized. Since ICA can decompose complex signals into their elemental components, it is possible to associate signal patterns critical for classification of cyber-attacks. Subsequent to that, for classification1 various models MLP, DT, NB were used. The results show that Z-score normalization and ICA have the better performance, especially combine with MLP model in accuracy, precision and recall than other case. This suggests that utilizing feature preprocessing and ML approaches together may offer a way of boosting the classification accuracy for cyber hacking tasks. Future research in cyber hacking breach classification should focus on areas that could enhance detection and response capabilities. One option may be to develop and leverage ML approaches, such as DL models that could improve accuracy and handle more complex threat behaviors. It may also be worth looking into utilizing real-time data streams with adaptive algorithms, which can rapidly adjust to new threats as they arise. Rather, improved interpretability of models could dramatically strengthen the speed and quality of answers through clearer understanding decision-breaking down frameworks.

References

1. Ch, R., Gadekallu, T.R., Abidi, M.H., Al-Ahmari, A.: Computational system to classify cyber crime offenses using machine learning. Sustainability **12**(10) (2020)
2. Veena, K., Meena, K., Kuppusamy, R., Teekaraman, Y., Angadi, R.V., Thelkar, A.R.: Cybercrime: ıdentification and prediction using machine learning techniques. Comput. Intell. Neurosci. (2022)
3. Chandrakala, T., Nirmala Sugirtha Rajini, S., Dharmarajan, K., Selvam, K.: Implementation of data mining and machine learning ın the concept of cybersecurity to overcome cyber attack. Turk. J. Comput. Math. Educ. **12**(12), 4561–4571 (2021)
4. Li, S., Zhao, S., Min, G., Qi, L., Liu, G.: Lightweight privacy-preserving scheme using homomorphic encryption in industrial Internet of Things. IEEE Internet Things J., 14542–14550 (2020)
5. Hulayyil, S.B., Li, S., Xu, L.: Machine-learning-based vulnerability detection and classification in ınternet of things device security. Electronics **12**(18) (2023)
6. Kumari, A., Prakash, P., Umadevi, M.: Prediction of data breaches using classification algorithms. In: 2021 Fifth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), Palladam, India, pp. 1049–1054 (2021)
7. NaifAlatawi, M., et al.: Cyber security against intrusion detection using ensemble-based approaches (2023). Article ID 8048311
8. Cremer, F.: Cyber risk and cybersecurity: a systematic review of data availability. Geneva Pap. Risk Insur. Issues Pract. **47**(3), 698–736 (2022)
9. Prathap, M.R., Nandhini, K.M., Vairavel, K.S., Suraj, M.V.: Detection of data breaching websites using machine learning. In: 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA), Coimbatore, India, pp. 1–6 (2021)
10. https://www.kaggle.com/datasets/teamincribo/cyber-security-attacks
11. Dasgupta, D., Akhtar, Z., Sen, S.: Machine learning in cybersecurity: a comprehensive survey. J. Defense Model. Simul. Appl. Methodol. Technol. **19**(1) (2020)

An Ensemble Method for Insider Threat Detection Based on User Activities Analysis Using Bi-LSTM and GA Optimization

A. Anju[1(✉)], Adline Freeda[2], Krithika Venket[3], Babymol Kurian[1,2,3], D. Sudha[1,2,3], and W. Mercy[1,2,3]

[1] Department of Information Technology, KCG College of Technology, Chennai, India
anju.it@lcgcollege.com, babymol.it@kcgcollege.com
[2] Department of Artificial Intelligence, KCG College of Technology, Chennai, India
adline@kcgcollege.com
[3] Department of CSE, KCG College of Technology, Chennai, India

Abstract. One of the main reasons why firms experience security breaches is insider threats. They are users or workers of an organization who carry out any malicious conduct with the intention of harming others. The majority of insider threat detection techniques currently in use depend on deep learning and machine learning techniques and have the following drawbacks: they need obvious feature engineering, which increases the number of false positives; they rely on established criteria or maintained patterns and fail to detect novel or unidentified threats they are highly computational and need a large amount of training data. For an improved client behavior-based insider threat identification system, this study proposes an integrated learning strategy to overcome the aforementioned limitations. An ensemble model Bidirectional long-short-term memory and uses a CNN, ANN with GA and Meta Learner also used to find the best solution by combining all models predictions. The first kernel selection for CNN is done using the fast global search method of the genetic algorithm.

Keywords: Bi-LSTM · CNN · GA · Insider Threats · Meta Learner

1 Introduction

The root cause of significant security lapses in companies is insider threats. They are generally divided into two categories: traitors and masqueraders, both are detrimental to any organization. They are workers or users of an organization who harm others by acting maliciously. There are two types of user behavior in an organization: malevolent and non-malicious. An insider threat is characterized as a user acting maliciously. The degree of harm caused is the main reason insider risks are regarded as one of the most challenging to counter. As per the findings internal threat survey report, 68% of firms are currently susceptible to insider threats. As such, it is imperative to detect insider risks with greater threat identification accuracy and a faster threat identification time. Designing an internal threat identification system is crucial for protecting a company's resources and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 80–87, 2025.
https://doi.org/10.1007/978-3-031-86305-9_7

reputation. The system should offer a proactive approach to recognizing and mitigating insider threats, guaranteeing a safe and resilient organizational environment, by utilizing sophisticated data analysis, machine learning, and real-time monitoring.

Insider threats are a serious danger to enterprises because they can result in lost intellectual property, financial loss, and reputational harm. Insiders, like staff, consultants, or collaborators, are the reason for insider threats, not external ones. These risks might be deliberate—like stealing confidential information—or inadvertent—like unintentionally leaking data. It is difficult to identify insider threats as harmful activity must be distinguished from the usual, lawful behavior of people you may trust. The majority of current detection of insider threats approaches are built on methods based on machine learning and deep learning.

Conventional machine learning techniques match user behavioral patterns (users' action sequences on a designated system inside an organization) using established rules stored signatures derived from quantitative user behavior data. Nevertheless, no particular collection of criteria can model and identify an insider danger because they cannot identify threats where the circumstances or characteristics are unknown, making them unable to identify novel, unidentified, or zero-day risks. In order to address this limitation, sophisticated machine learning techniques were implemented. They don't rely on preset rules or saved signatures, thus they can identify novel or unknown threats, but they do involve intentional feature engineering, which leads to an increased quantity of false positives. Therefore, in order to overcome this limitation, deep learning techniques were eventually given more attention [1]. While these techniques produce effective and high-quality results, their high computing costs and extensive training data requirements make them unsuitable for widespread use. In order to improve on current techniques, this paper provides an innovative mixed learning approach for insider threat identification. The suggested approach detects insider risks by utilizing numerous modules. Preprocessing of data entails reduction via locally linear embedding, transformation with a preference matrix, and cleaning. The preferred matrix assists in creating the best dataset while maintaining the dataset's nonlinearity, taking into account a number of factors like accuracy and computing time. These actions increase classifier accuracy while lowering classification overhead.

The suggested approach detects insider risks by utilizing numerous modules. Preprocessing of data entails reduction via locally linear embedding, transformation with a preference matrix, and cleaning. The method helps to reduce dimensionality while keeping dataset unpredictability, and the chosen matrices aids in the preparation of the ideal dataset depending on a variety of parameters such as accuracy, precision, processing time, and so on. These actions increase classifier accuracy while lowering classification overhead. Then, by comparing and matching the present action with the previous action, bi-directional long short-term memory (Bi-LSTM) and CNN is utilized to extract behavioral-based information from the user [3]. It identifies insider threats without the need for signature storage or set restrictions, allowing it to identify suspicious activity beforehand.

Bi-LSTM lowers error rates by memorizing occurrence patterns for both the past and the future, or substantially more effectively than a conventional LSTM, without necessitating specific engineering features. The similarity of feature vectors is evaluated

using an artificial neural network (ANN) and three distance-measuring measures. The validation between the two constraints, user-by-user with action-by-action comparison, relies on the observation that anomalies are identified by distinct user behaviors [16]. We evaluated feature similarity using the Euclidean distance, cosine similarity, and standard deviation. Identifying all of the essential characteristics from the obtained set of characteristics decreases the number of false positives.

2 Related Work

The reconsideration of data collection could lead to better results. The freshly constructed preference matrix aids in identifying the best normalizing strategy for a particular set of data. Produce quickly, precisely, and simply. Different data collection designs may yield different results. A simple design could cost us time or accuracy [13]. Assume for the moment that we select the most accurate and straightforward data set design. If the data collection is non-linear, this design may result in an extremely long computing time, which is undesirable. By assigning weights to each criterion, the augmented preference matrix aids in selecting the best data set design. Preferred factors are listed in columns, and data set designs are listed in rows. There are n factors total. There must be a minimum of one behavioral attribute [4]. Two user activities may share the same priority if their numbers are equal. In terms of computation time, a rapid computing-time data set design has priority 1 (the lowest priority). The greatest precision, measured by means of accuracy factor, is 1. Next, the highest accuracy level is 2, and so on. The preference matrix's last column contains the overall number of preferences.

The last column of every row totals up all of the criteria in that row. The chosen data set for the specific row has the most suitable we could offer because the final segment of the matrix has the highest significance. This gathering of data has the best design when all three factors are taken into account. If convenience or processing speed are not as crucial as accuracy, then another data collection method can be preferable. Columns show precise data mining techniques, and rows show various data set configurations. A data set design with several highly accurate approaches is optimal. The optimal data set summary has the lowest priorities. The best designed data set is the least summation data set [5].

The prevalence of deep learning and machine learning algorithms has greatly enhanced insider threat detection techniques. This section presents the most modern strategies for detecting insider threats using machine learning, statistical analysis, and deep learning. Below is a discussion of each category. Nonetheless, a number of variables influence how well these techniques work. These include, among other challenges, a lack of preprocessing, insufficient information, superfluous or noisy information, large-scale characteristics, the need for particular characteristic generation, nonlinear interactions between attributes, and a dynamic environment [13].

3 Proposed Methodology

A. Data Preprocessing

A raw data set may contain noise in many different ways, including null values, duplicate or missing numbers, improper data formats, etc. The recommended method states that in order to perform hybrid learning, data preprocessing comprises transforming user information into a structure that the system can understand, as follows: The proposed methodology is shown below in Fig. 1:

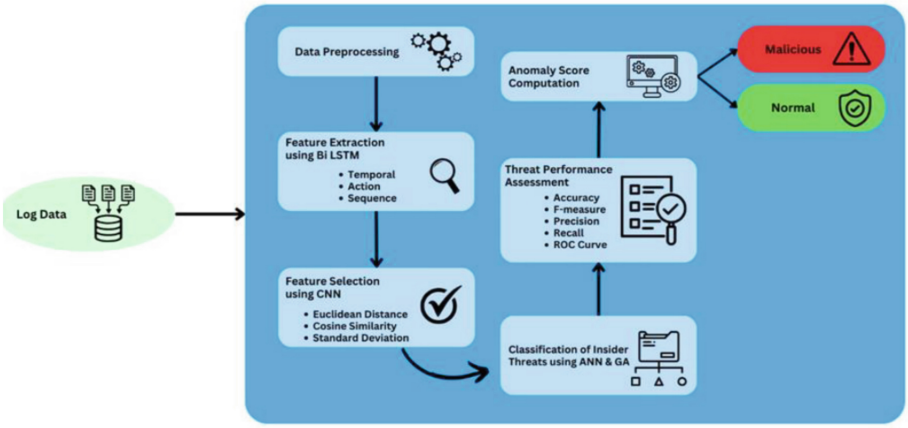


Fig. 1. Overview of the proposed methodology

1. Data cleaning: At this stage, lost data is restored, and arrays with missing values are filled in. We identified the missing numerical attributes and substituted "0" for the missing values. The category features that were absent were categorized as "Missing." We also searched the data set for redundant user entries [6]
2. Data Transformation: During this process, just the numerical properties of the dataset are normalized to a predefined limit. A decision matrix is utilized in order to standardize data as opposed to the min-max technique for normalization [1].

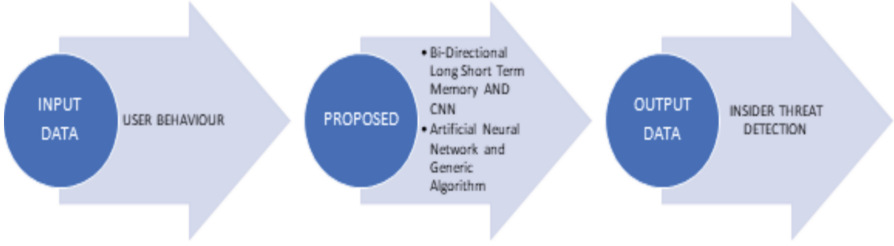


Fig. 2. Hybrid User Behavior based Classification

B. **Feature Extraction**

The capability of LSTM can identify lengthy time-series correlations makes it a feasible choice for conversational modeling. The bi-directional LSTM proposed in this research uses the optimizer Adam to learn hourly user actions, associated action sequences, and subsequently their actions in language. Bi-LSTM is the most effective method for resolving the standard LSTM disadvantage. For long time series-based action sequence prediction, the suggested Bi-LSTM framework can learn behavioral features of the user and recover the feature order in all directions (ahead - past toward future times / backwards - future towards past). For instance, each user is allocated a system that stays on the company's server and is aware of their regular login and logout times [15] (Fig. 2).

4 Classification

This section classifies users as either malicious or normal based on past and current activity sequence. BiLSTM + CNN + ANN + GA—are combined in the suggested classification. The best applications as a classification algorithm are dichotomous class problems (positive and negative). The classification process takes into account the following stages: (1) Choose a kernel function; 2) selecting a kernel parameter (a, β); (3) optimizing quadratic programming; and (4) creating classification functions from vector supports. However, CNN's first kernel selection is a very difficult operation that results in low accuracy. Thus, in this study, we first select the best-fit kernel for ANN using GA, and then we detect insider threats using a standard ANN.

Without data preparation, the learning algorithm operates more slowly and inefficiently. In this work, data preparation methods are applied before raw data is entered for feature extraction. Raw data is reduced, cleaned, and transformed in order to do this. The data is produced by decreasing the input via locally linear embedding, cleaning, and processing using a preferred matrix [5]. The preferred grid can be used to lower a dataset's dimensionality while preserving its inherent non-linearity, and it can assist in creating the perfect dataset depending on a number of factors, such as accuracy and computation speed, among others. These techniques drastically cut down on classification overhead [14].

For feature extraction, unidirectional LSTM preserves data in a single direction (forward). In this work, the user's action sequence was extracted and stored in both forward and backward orientations using bidirectional LSTM. In order to successfully extract features, this allowed us to save data from the past, transfer it to the future, and then return it to the past. The recommended technique to threat identification is not dependent on preset rules or stored signatures. Malicious users are instead found through analyzing their actions using log data, which allows for the considerably earlier detection of suspicious activities. By collecting each user's whole collection of capabilities (users behavior-action sequence) without the need for specific feature design, it lowers the number of false positives.

5 Results and Discussions

The data set used here is publically available CERT r4.2. This dataset is a compilation of employees' regular and irregular activities throughout an 15-month period within the company. Since just 65 of the 950 users in the sample are malevolent individuals or insider threats, the dataset is severely unbalanced. 900 of the 950 users are actual user. Thus, insider threat incidents in our research represent only 0.34% and 0.27% of the pre-processed data on a weekly and daily basis, respectively. As a result, the suggested approach's outcomes using the premise and that compared to regular users, there will be significantly less malevolent users in the real world. Furthermore, we ensure that the same imbalanced dataset used in this study is used to confirm the outcomes of the contrasted cutting-edge insider threat detection algorithms. Each line segment (entry) in this dataset has the following attributes, all of which are produced in formats that are comprehensible to humans: Action details include one of the following: the actions the user took on the system they were assigned: Log Activity (starting, stopping); File (cutting, copying, pasting); Email (sending, receiving); HTTP (viewing URLs); Device (connecting, disconnecting removable devices). A total of sixteen different activity types—device, file, HTTP, email, and logon-logoff—were taken into consideration [3].

Five metrics are used to evaluate the effectiveness of the suggested method: AUC (Area Under Curve), recall, accuracy, precision and F-measure. It is essential to perform data preprocessing correctly prior training the model since it has a direct impact on the success rate. Several data preparation techniques, including reduction, transformation, and cleaning, were used in this investigation. Then, for feature extraction, Bi-LSTM is employed. Because of its long-term memory for temporal dependencies, it is used. There's no direct correlation between the CSV files in the collection since the insider threat data that is provided is nonlinear. As such, we extracted features from the respective.CSV files produced with the Bi-LSTM.

Additionally, because that can be difficult to categorize high-dimensional non-linear data, we only selected the best or almost finest qualities from the features that were extracted set. Using the complete feature set could have resulted in inferior classification results. An ANN condenses the obtained set of features by evaluating both the requirements of user-by-user and action-by-action matching for improved precision,

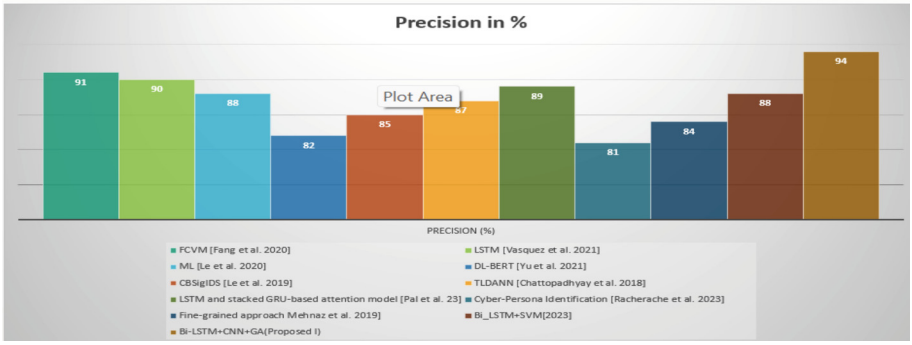


Fig. 3. Precision Comparison with proposed system

because harmful conduct of a user can be detected by comparing his action sequences to those of other users. Thus, this work offers both instance-based and user-based evaluations. A grid search strategy is also used to weak the hyperparameters of the ANN in order to prevent overfitting of the model. All possible combinations of the GA's parameters such as population, mutation rate, number of generations, crossover percentage, and percentage of mutations on the population can be adjusted using grid search (Fig. 3).

6 Conclusion and Future Work

The suggested approach uses user behavior analysis to identify insider threats and malevolent people. The different models have undergone training, validation and testing phases. A meta learner Logistic regression algorithm ensemble learning approach to aggregate predictions from multiple (models) and produce a final prediction. Each model eliminates the hourly activity sequences from every user's everyday schedule. Relevant feature selection is done using an artificial neural network. Three distance measurements are used in this to verify the dual restrictions of the action-by-action and user-by-user matching. To maximize its performance, CNN's first kernel is chosen using a genetic approach. Due to the significant skewness of the proposed dataset, only the majority class—normal user behavior—rather than the minority class—malicious user conduct—is taught. This causes the results to be interpreted incorrectly. Given this, future dataset balancing will lessen the likelihood of over fitting a model with accurate result analysis. Furthermore, examining additional user behavioral data sets, such the contents of emails sent and received, files accessed, and URLs visited, can yield valuable insights that facilitate the distinction between benign and malicious activity.

References

1. Singh, M., Mehtre, B.M., Sangeetha, S., et al.: User behaviour based insider threat detection using a hybrid learning approach. J. Ambient Intell. Hum. Comput **14**, 4573–4593 (2023)
2. Anju, A., Shalini, K., Ravikumar, H., Saranya, P., Krishnamurthy, M.: Detection of insider threats using deep learning. In: 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN), Salem, India, pp. 264–269 (2023). https://doi.org/10.1109/ICPCSN58827.2023.00050
3. Liu, L., De Vel, O., Han, Q.L., Zhang, J., Xiang, Y.: Detecting and preventing cyber insider threats: a survey. IEEE Commun. Surv. Tut. **20**(2), 1397–1417 (2018)
4. Meng, F., Lou, F., Fu, Y., Tian, Z.: Deep learning based attribute classification insider threat detection for data security. In: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), pp 576–581. IEEE (2018)
5. Anju, A., Abhilash, M.R., Balaji, V., Sayee Darshan, J.K.: Third vision for safeguarding women with a live surrounding update using deep-learning for computer vision. In: 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), Solan, India (2021)
6. Lv, B., Wang, D., Wang, Y., Lv, Q., Lu, D.: A hybrid model based on multi-dimensional features for insider threat detection. In: Chellappan, S., Cheng, W., Li, W. (eds.) International Conference on Wireless Algorithms, Systems, and Applications, vol. 10874, pp. 333–344. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94268-1_28

7. Almehmadi, A.: Micromovement behavior as an intention detection measurement for preventing insider threats. IEEE Access **6**, 40626–40637 (2018)
8. Anju, A., Barath, M., Maheshwaran, R., Vignesh, K., Sr, K.K.: Sentimental analysis for E-commerce website. In: 2022 10th International Conference on Emerging Trends in Engineering and Technology (2022)
9. Chattopadhyay, P., Wang, L., Tan, Y.-P.: Scenario-based insider threat detection from cyber activities. IEEE Trans. Comput. Soc. Syst. **5**(3), 660–675 (2018)
10. Böse, B., Avasarala, B., Tirthapura, S., Chung, Y.Y., Steiner, D.: Detecting insider threats using radish: a system for real-time anomaly detection in heterogeneous data streams. IEEE Syst. J. **11**(2), 471–482 (2017)
11. Anju, A., Krishnamurthy, M.: M-EOS: modified-equilibrium optimization-based stacked CNN for insider threat detection. Wireless Netw. **30**, 2819–2838 (2024)
12. Lansky, J., et al.: Deep learning-based intrusion detection systems: a systematic review. IEEE Access **9**(1), 101574–101599 (2021)
13. Le, D.C., Zincir-Heywood, N.: Exploring anomalous behaviour detection and classification for insider threat identification. Int. J. Network Manage **31**(4), 2109–2119 (2021)
14. Li, B., Wu, Y., Song, J., Lu, R., Li, T., Zhao, L.: DeepFed: federated deep learning for intrusion detection in industrial cyber–physical systems. IEEE Trans. Industr. Inf. **17**(8), 5615–5624 (2020)
15. Logeswari, G., Bose, S., Anitha, T.J.I.A.: An intrusion detection system for SDN using machine learning. Intell. Autom. Soft Comput. **35**(1), 867–880 (2023)
16. Lv, H., Ding, Y.: A hybrid intrusion detection system with K-means and CNN+ LSTM. EAI Endorsed Trans. Scalable Inf. Syst. **11**(6), 1–10 (2024)

Systemization of Knowledge: Privacy Preservation for Decentralized Non-financial Applications

Purnima Ahirao(✉) and Deepti Reddy

Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS, Mumbai, India
purnimaahirao@somaiya.edu, deepti.reddy@nmims.edu

Abstract. The digital ecosystem is used extensively by people worldwide for various services such as social media, banking, healthcare, insurance, etc. Every coin has two sides similarly digital usage also has its advantages and disadvantages from the user's perspective. One of the disadvantages is the sharing of personal and sensitive data over the digital medium without understanding the hazards. This personal identification information (PII) is used not only for providing personalized services to the users but also for profiling the users for the gain of the service providers. The importance of user rights over their data is easily neglected. Privacy Preservation plays an important role in the digital ecosystem to control the security and privacy of the users PII. This includes achieving User Anonymity, Data Confidentiality and Access Control. Researchers have provided various methods and solutions to ensure privacy and security of digital data using centralized methods including cloud services. Security challenges exist in such approaches due to their centralized nature. Eventually, researchers are exploring the decentralized approach using blockchain technology. Blockchain technology paves the way for many applications that need a distributed and trustful way of data management. This paper presents a systematic review of research work done to satisfy the privacy preservation challenges in decentralized, non-financial applications such as Healthcare, Social media, Identity management, etc., on a public blockchain i.e. Ethereum, and further identifies the gaps amongst them. The authors conclude by suggesting a combined approach to solve the existing gaps for efficient privacy protection of digital data on Dapps.

Keywords: Privacy Preservation · Decentralized Application · Ring Signature · Homomorphic Encryption

1 Introduction

The world is becoming more and more data-centric and data-driven. Data has become the new fuel that is gathered and analyzed for economic growth and innovation. Although a data-driven culture benefits all, user privacy is becoming more of a concern. The service providers collect data in many forms for providing online services. Some of the

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 88–104, 2025.
https://doi.org/10.1007/978-3-031-86305-9_8

data collected can be categorized as Personal Identifiable Information (PII). In paper [1], PII is considered as to uniquely identify an individual in the digital ecosystem. As more and more people join the digital revolution, security and privacy preservation is the basic need that requires some robust solution. Authors in [2] mention that the traditional way of data storage often adopts centralized architecture and secures it with different privacy mechanisms such as cryptography, digital signature, etc. This is prone to trust and security problems. Security risks are always high due to the dependency on the third party in centralized architecture. Sharing and storing data in a secure manner is important. Also providing fine-grained access control to multiple entities or stakeholders is an important aspect that remains a challenge in a centralized architecture. To overcome these challenges, a decentralized approach i.e. Blockchain is being explored and used for development of secured and privacy enriched applications. Paper [3, 4] discusses blockchain to be treated as the most secure, trusted, and immutable system to store information due to its cryptographically secured and consensus-based mechanism. Blockchain based applications are gaining popularity in many sectors that include multiple stakeholders and data integrity provision. Authors in [5] describe that Distributed Applications (Dapps) are being developed for various applications using the public blockchain Ethereum such as healthcare management, supply chain management, inventory management, government services etc. Data storing, management and access control are possible due to its ledger based and cryptographically driven fundamentals. Organizations including government's agencies are experimenting with this technology to deliver efficient digital services. In the paper [6], it is stated that blockchain architecture lacks privacy as the users are pseudo anonymous, the data is publicly viewable and access control needs to be designed. Researchers are striving hard to build privacy features on top of blockchain architecture to make it appropriate for robust Dapps serving data-rich applications such as healthcare and government services. Authors in [7] suggest advanced cryptographic techniques such as zero-knowledge proof, homomorphic encryption, ring signature, differential equation, etc. are being used to build privacy parameters for Dapp. This systemization of knowledge provides a framework to understand existing research work in adding privacy to blockchain applications using public blockchain Ethereum. The framework initially categories the different cryptographic primitives available for achieving privacy preservation parameters on top of blockchain. Secondly, it critically reviews the current state of art and identifies gaps. This study further addresses the gaps identified by proposing an integrated solution using Ring Signature, homomorphic encryption and smart contract based access control for achieving each of the privacy parameters and integrating security techniques over the enriching efficacy of the blockchain technology.

2 Literature Review

2.1 Blockchain: An Undisrupted Technology

A blockchain is a distributed ledger that preserves an unchangeable record of all network transactions. Research work in [1–6] ensures that block-chain-based solutions can deal with the problems of data management, data sharing, and data storage in a more intrinsic manner for Dapps It can offer a reliable environment for sharing and storing data,

essential for any digital business because of its decentralized nature and difficulty in being tampered with. Blockchain, a peer-to-peer (P2P) network, is distinguished by its high-level security performance and immense features like verification, authentication, decentralization, and synchronization. It also avoids the need for a centralized architecture. This ensures a more robust model to address the integrity and privacy of digital data with user-centric access when compared with existing security and privacy mechanisms.

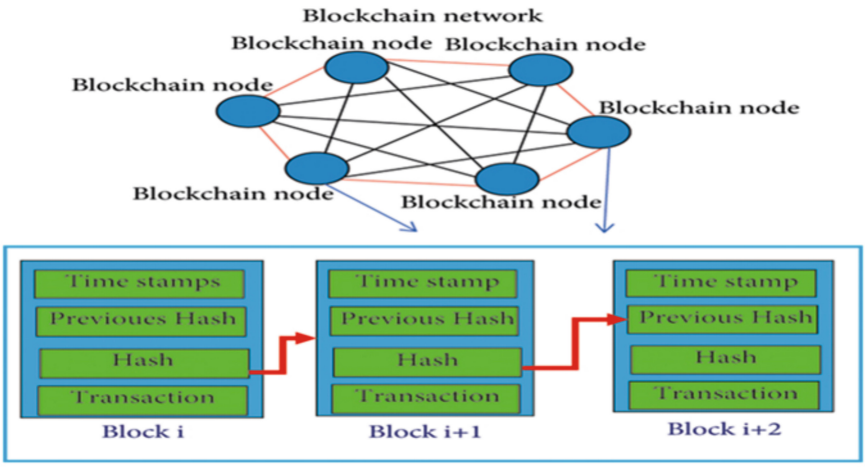


Fig. 1. Blockchain Architecture

From the paper [33], we understand that blockchain uses cryptography to support transaction integrity using digital signature for building trust in the network. As shown in Fig. 1, transactions are timestamped and included in a block, the blocks are then subjected to verification and validation before being added to the blockchain. Cryptographic hash of the transactions is derived by solving a proof of work equation including previous blockhash, merkletree and one-time nonce. This operation makes the blockchain secure and immutable as tampering of any one transaction changes the whole block and subsequently the rest of the chain of blocks. Authors [22] believe that the intrinsic properties of blockchain technology promote security by addressing the information security triad of Availability, Integrity, and Security (CIA). Blockchain employs digital signatures and encryption to promote transaction integrity and network trust. Blockchain capabilities include audit trails and traceability, the use of consensus mechanisms to commit transactions, and transaction immutability. Similarly, authors in [5–7] believe that by removing single points of failure, blockchain technology can increase availability. The ledger and smart contract execution are decentralized; if a node fails, the network can still function and the data remains readable. The author also agrees that blockchain technology has revolutionized several operations, including asset ledgers, financial transactions, global supply chain management, decentralized information exchange, and healthcare data management.

2.2 Challenges in Blockchain Applications

Though the blockchain technology has some built-in security mechanisms by design, the data privacy concerns such as user anonymity, data confidentiality and fine-grained access control remains a challenge. Scholars [21] have examined these underlying challenges faced by blockchain users and regulators globally. They agree about the different applications where blockchain can be used as a ledger but they argue that when huge volume of date gets captured and stored there is a need for data protection and privacy that is also a regulatory requirement and needs to have compliance such as GDPR. The challenge is the conflict with the European data regulation framework i.e. in GDPR, it clearly states the access rights of the users data including "right to be forgotten" (RTBF) rights of the users. The immutable nature of data on the blockchain is in conflict with this right even though all other access rights can be achieved by using smart contracts that follow a Turing complete structure. Other arenas [19] that scholars and researchers are continuously working for improvising are the mechanisms to provide user anonymity, data confidentiality and access control. Blockchain architecture lacks privacy by design fundamentals. Though it provides integrity and immutability along with all other advantages of decentralized architecture, it does not have privacy by default.

2.3 Public Blockchain: The Privacy Factors [24, 26]

User Anonymity: Bitcoin and Ethereum follow pseudo-anonymity i.e. when you make a transaction with Bitcoin or Ethereum, your identity is not directly tied to your address. Instead, transactions are recorded using cryptographic addresses that are strings of characters unique to each user through the use of a wallet. This indicates that your real identity is not immediately obvious from your wallet account. But as the name suggests, it does not provide full anonymity to the user and the identity can be revealed by observing the patterns. While your identity isn't directly linked to your wallet address, patterns in your transactions, along with other information, can sometimes be used to infer your identity. Hence linkability is an issue in using wallets and correspondingly public blockchain.

Data Confidentiality: All transactions are recorded on a public ledger called the blockchain. This ledger is accessible to anyone, allowing them to view all transaction history.

Access Control: Public blockchain Ethereum or Bitcoin does not have any way to provide access control or authorization in its default state.

This leads to the conclusion that public blockchain has several privacy concerns such as if someone knows the wallet address they can link it to real identity and they can track all transactions on the blockchain. Data is freely viewable and no access or authorization available in Public Blockchain. Hence public blockchain cannot be used for providing privacy preservation to digital data in its default state of implementation.

2.4 Public Blockchain: Privacy Preservation Mechanism

Researchers [6, 33] have implemented advanced cryptographic techniques to add a layer of privacy to the blockchain application to make it a robust solution in varied sectors.

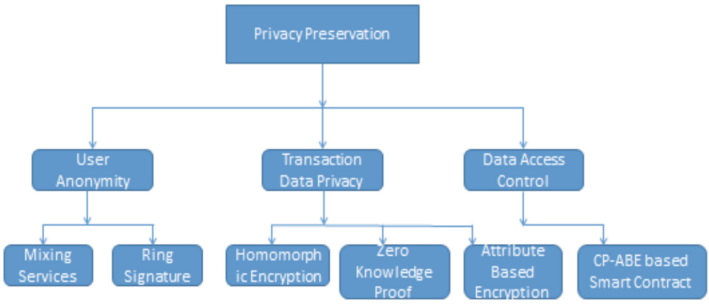


Fig. 2. Privacy Preservation Parameters and Techniques

Figure 2 shows the different factors of privacy preservation and their corresponding advanced cryptographic mechanism to be added as a privacy layer on top of block chain's inherent capabilities of authentication, integrity, availability etc.

Improving User Anonymity: User Anonymity can be Improved by Achieving Unlinkability [23–25]

Mixing Services: The purpose of mixing services, sometimes known as tumblers, was to keep user addresses separate. The user's coins are mixed at random. As a result, their coin ownership remains obfuscated from other peers. In 2013, CoinJoin was used as a substitute anonymization technique for Bitcoin transactions. The concept of shared payment serves as its motivation. It is a mechanism that makes it challenging to link specific inputs to specific outputs by combining numerous users' transactions into a single transaction. Similarly, coin mixing services mix your bitcoins with those of other users, obscuring the transaction trail. These services help enhance privacy by breaking the link between your Bitcoin address and your transactions. Mixcoin provides anonymous payment in Bitcoin and bitcoin-like cryptocurrencies.

Privacy-Focused Wallets: There are wallets specifically designed to prioritize privacy. These wallets may incorporate features such as coin selection algorithms that optimize transaction outputs for privacy, and integration with privacy-enhancing technologies like CoinJoin.

Stealth Addresses: Stealth addresses generate unique addresses for each transaction, making it harder to link transactions to a specific user. When you receive bitcoins, they are sent to a one-time address derived from your public key, ensuring that your actual Bitcoin address remains private.

Utilize Privacy Coins: Some cryptocurrencies, known as privacy coins, are designed to prioritize anonymity and privacy by default. Examples include Monero, Zcash, and Dash. By using these cryptocurrencies for transactions, you can benefit from enhanced privacy features built into their protocols.

Ring Signatures: Authors in [24, 26] describe ring signature in cryptography as a method to obfuscate identities of the parties involved in a transaction and create a type of digital signature. It uses any of the public key cryptography algorithms such as ECC, RSA, or bilinear mapping as a mechanism that mixes the identities of sender and recipient in a transaction. Ring signatures allow a group of signers to produce a

signature on behalf of a message without revealing which member of the group actually produced the signature. This can be used to obfuscate the sender's identity to provide transaction privacy. The signature technique that makes use of the ring-like structure is where the phrase "ring signature" comes from. If the number of ring members is n, then the probability that an adversary may successfully guess a real sender of a transaction is 1/n. CryptoNote uses ring signatures to conceal the relationship between the sender's transaction addresses. To put it more succinctly, CryptoNote creates the sender's public key using several different keys, making it impossible to figure out who transmitted or signed the transaction. Another real-time application is Monero, that provides anonymity and transaction privacy using stealth addresses, ring signatures, and ring ct.

Data Confidentiality

Homomorphic encryption: Paper [28] describes Homomorphic encryption (HE) as a powerful cryptography technique for data confidentiality. It can carry out specific computations directly on ciphertext and guaranteeing that actions taken on the encrypted data will produce outcomes that are identical to those taken on the plaintext when the computed results are decrypted. Without materially altering the characteristics of the blockchain, data can be stored and processed using homomorphic encryption algorithms. Using this privacy issues with public blockchain are addressed and the blockchain's data is guaranteed to be encrypted. Ethereum smart contracts can be designed to implement homomorphic encryption. In Healthcare applications, Sensitive health data can be encrypted using homomorphic encryption. Even when processed, the data remains encrypted, preserving privacy. Only authorized parties can decrypt and access the original information. Homomorphic encryption significantly impacts e-voting systems by enhancing privacy, verifiability, and security. It allows votes to be recorded and stored in a blockchain network. Secrecy of Intermediate Results can be achieved such as vote tallies remain secret; votes are counted in their encrypted state and no of voters also can be counted without decrypting the exact count. At the same time trustworthiness of final results can be achieved by directly verifying the vote count in its encrypted form. Homomorphic encryption Encrypted data remains secure, and the integrity of the voting process is maintained.

Encrypted Transactions: Users can encrypt transaction data using cryptographic techniques, such as symmetric or asymmetric encryption, to ensure that only the intended recipient can access and decrypt the transaction details.

Zero-Knowledge Proof: Transaction Obfuscation/Verification and user authentication. The basic idea in [18] is to generate a proof, formulated to verify that a pro-gram executed with some input known by the initiator only and produce some publicly known output with no disclosure of any other information. Here, a certifier can demonstrate to a verifier that a particular assertion is true without giving the verifier any relevant information. As a variation on zero-knowledge proofs, it is demonstrated that computational zero-knowledge can be obtained using the non-interactive variant of zero-knowledge proofs, or NIZK. Here if the certifier and the verifier have the same reference string, computational zero-knowledge can be achieved without any interaction between the two. All account balances are encrypted and kept in a public blockchain. Without revealing the account balance, a user can effortlessly demonstrate to another user that he has

enough money for the transfer using zero-knowledge proofs. This will ensure transaction obfuscation without compromising user authenticity. Zcash uses zk-SNARKs to verify transactions while protecting users' privacy. The Ethereum contract language was recently improved by the Zcash group to efficiently allow zk-SNARK proofs verification. They particularly adapted a snark-verify precompile (similar to an op-code) to a clone of "Parity," that verifies generic proofs using lib-snark. Additionally, they enforced a novel coin mixing contract using the new zk-SNARK verifier. Tornado Cash is another real-time example; it provides decentralized and non-custodial service. The users can carry out private transactions on the Ethereum Blockchain. It uses Zero-knowledge proof for transaction obfuscation also ensuring user authentication due to the inherent capability of zero-knowledge proof.

Access Control: Attribute Based Encryption
Fine-grained access control of encrypted data is made possible by attribute-based encryption, a generalization of public-key encryption that uses authorization policies. A user's secret key and the ciphertext depend on various variables, such as their email address, nationality, or kind of subscription. In a system like this, a ciphertext can only be decrypted if the user key's set of attributes corresponds with the ciphertext's attributes.

Privacy Techniques and Their Use Cases
Table 1 depicts the privacy preservation mechanism used in the financial applications such as Monero, Zcash, Zerocash, Tornado, Cryptonote etc., over the public blockchains Bitcoin and Ethereum. Continuous Research is going on, to make more robust in terms of achieving overall privacy preservation.

Table 1. Privacy Preservation Techniques and their use cases

Approach	Use Cases
Mixing Services	CoinJoin proposed in 2013 serves as an alternative anonymization approach for transactions in bitcoin Mixcoin proposed by Bonneau et al. in 2014, provides user anonymity for payment transaction in bitcoin and other cryptocurrencies
Ring Signature	CryptoNote, an application using ring signature. It adopts this approach to hide the linkage between the sender's addresses and the corresponding transactions Monero applies ring signature, ring ct and stealth address thus ensuring anonymity and transaction privacy
Zero-Knowledge proof	Research has used zero-knowledge proof for Transaction verification. The content is verified without revealing the actual data NIZK has been used for authentication and. Zerocash and Zcash uses zk-snarks Tornado Cash is another real-time example; it provides decentralized and non-custodial service. The users can carry out private transactions on Ethereum Blockchain. It uses Zero-knowledge proof for transaction obfuscation also ensuring user authentication due to inherent capability of zero-knowledge proof

2.5 Blockchain Use Cases: Security and Privacy

Many Non-Financial applications are being developed using Blockchain Technology such as healthcare applications, social media applications, identity management; government services such as Land registration system, voting system, these applications require data management in terms of data storing, processing and access control. The privacy preservation factors and mechanism discussed above are necessary and applied over these applications too.

Healthcare Applications: As discussed in [10], offering enhanced security, increased transparency, and patient-centered interoperability, blockchain technology can help in better management of healthcare record. Blockchain is one such novel technology that could have a role in improving interoperability. Citing the following reasons eminent researchers [12] are considering blockchain based application as a solution for patient-centric electronic healthcare system.

Privacy and Security of Electronic Health Record: As per authors in [11], The vital security and privacy requirements in e-health systems are.

- Data integrity-No alteration in the data stored for use in an application.
- Data confidentiality- Encrypting the data in the transit or while stored is the only way to ensure data confidentiality
- Authenticity- This feature ensures that only valid users can access the data
- Accountability-It is the way to ensure that someone is responsible for the information
- Audit- This requirement needs to track the use of the health record while moving in and out of the functions
- Non-repudiation- This is important feature of any digital data transfer, it ensures that identity of the sender is authenticated by the use of digital signature during and after the transaction.
- Anonymity- Data privacy needs to hide the identity of the user while making transactions.

While developing a patient centric health care application on a decentralized architecture, it is necessary to achieve all the above-mentioned privacy and security of the patient health record. Data needs to be stored, managed, and shared using blockchain. In paper [2] Attribute based encryption is proposed for ensuring privacy of data as there is no in-built mechanism in blockchain for privacy. Authentication can be achieved using Non-Interactive Zero-knowledge proof. Use of IPFS for distributed storage of data is suggested as the data collected by a mHealth system is extremely sensitive, and some security and privacy requirements are essential; the authors [3] propose a blockchain based framework for providing access to patient's data. Smart contracts are designed in an Ethereum-based blockchain for high-end access control, obfuscation of data, and advanced cryptographic techniques in this framework. Authorized Access, tracking and secure transfer is provided. Encryption and authorization is used throughout. The authors in [5] presents a blockchain based privacy preservation framework for securing EHR data. To guarantee the confidentiality, integrity, and availability of EHR, the suggested system combines cutting-edge cryptographic approaches with the immutability and decentralized nature of blockchain. The Inter-Planetary File System (IPFS), that houses the EHR data, is encrypted using a hybrid cryptographic technique.

Furthermore, the research [4] designs a novel smart contract-based patient-centric access control that safeguards patient data privacy by using a blockchain-based SHA-256 hashing method. To accomplish secure EHR sharing and privacy preservation, this research uses a hybrid cryptographic technique that combines the advanced encryption standard (AES), elliptic curve digital signature algorithm (ECDSA), and elliptic curve integrated encryption scheme (ECIES).

Social media: Many researchers [15, 16] have seen blockchain technology as a solution to control social media providers' data exploitation. Numerous research institutions are exploring decentralized solutions to provide a solution to the OSN user conundrum of privacy invasion by outside parties and hacks. Data Privacy ensures that users should be in charge of their access permissions and be able to decide whether to share their sensitive and personally identifiable information. As soon as the user expresses the necessary concern, the service provider can then utilize the data without a doubt. This leads to the need of developing a robust ecosystem that helps users to enjoy the required claim and ownership of the data and at the same time partake in the social media network's profits. Authors in [17] mention that Blockchain-enabled social media (BSM) is not a new concept. Ideas like distributed online social networks (DOSNs), decentralized autonomous online communities (DAOCs) based on blockchain technology, and decentralized online social networks (BSNs) can all be considered the early stages of BSM and lay the groundwork for its eventual rise to prominence. Decentralization, that embodies the distinct nature of its organizational structure and governance, is the primary characteristic of this new breed of blockchain-enabled company. Additionally, autonomy is the fundamental operating mechanism of the businesses, and it is automatically handled via established online norms and protocols. In paper [34] Smart contract for access control that performs transactions in the form of record of ownership of data items and number of shares that can be made on the blockchain is designed for access rights of user's data. The application is built on Ethereum.

Identity Verification: Paper [18] discusses SecureKey leverages blockchain to implement an identity verification network in Canada. When making a purchase of goods or services, a consumer can use the network to authenticate their identity. The blockchain network distributes an individual's identification attributes rather than concentrating them in one location. Here anonymity is provided by implementing "triple blind" technology, wherein the identity attribute provider is blind to whom they offer the attribute, the consumer is blind to the source of the attribute, and SecureKey is blind to the attribute exchange. For example, "B3i" is a consortium dedicated to establishing trading platforms across the whole insurance value chain developed using blockchain technologies. Every B3i member company has a private ledger where it keeps track of the information specific to its insurance contracts. Subsequently, every organization takes part in two common ledgers: Common and agreed-upon master data, such as public corporate information and common contract stipulations, contained in the shared master data ledger. Consent regarding the state of the private organization ledgers is communicated and stored in a cryptographically secured shared communication ledger. The communication's destination and substance are only visible to counterparties.

2.6 Integrating the Benefits of Advanced Cryptographic Techniques

Authors in [9] found that Ring Signatures and Zero-Knowledge Proofs can satisfy privacy preservation features to some extent without breaking the chain. The approach uses Fast High-dimensional Attribute Domain-based Message Encryption (HAD-FME) is proposed for data security. Smart contracts are built in conjunction with an Attribute Revocation Mechanism Based on Sentry Mode (SM-ARM). These two approaches are integrated to provide high-end privacy and security of transaction data. This paper [10] proposes a data access control scheme combining attribute-based encryption and blockchain that combines a symmetric encryption algorithm and attribute-based encryption algorithm to realize the fine-grained access control of data owners and protect the privacy protection rights of data owners. At the same time, this paper combines blockchain technology with distributed storage, and only stores the hash value of the data and the hash value of the ciphertext location, access control policy, and other information on the blockchain, that solves the storage bottleneck of blockchain system.

Table 2. Privacy Preservation Achieved in existing use cases

Paper title	Use case	Blockchain framework used	Privacy Parameters
Preserving Privacy in Mobile Health Systems Using Non-Interactive Zero-Knowledge Proof and Blockchain [2]	Mobile Health	Ethereum blockchain and Rinkeby network for testing and development IPFS library and IPFS client	Access control: Smart contract for access control is deployed on blockchain Data Confidentiality: ABE is used for confidentiality User anonymity: not mentioned
Privacy-preserving Framework for Access Control and Interoperability of Electronic Health Records Using Blockchain Technology [3]	Patients health record	Ethereum permissioned blockchain	Access Control: Smart contracts are used for access control. Has a separate storage that is not secured Data Confidentiality: Not mentioned User anonymity: not mentioned

(*continued*)

Table 2. (*continued*)

Paper title	Use case	Blockchain framework used	Privacy Parameters
Blockchain-Based Framework for Privacy Preservation and Securing EHR with Patient-Centric Access Control [4]	Health data	Public Ethereum blockchain Ganache and sepolia are used for testing	Data confidentiality: IPFS used for data storage. Data is stored in encrypted form Access control: Patient centric access control using smart contract
Privacy-Preserving Blockchain Framework Based on Ring Signatures (RSs) and Zero-Knowledge Proofs(ZKPs) [28]	No use case	Hyperledger Fabric permissioned Consensus is through proof of authority Nodes needs permission	User anonymity: Zero-knowledge proof and ring signature is used for verification and validation but the authors are not able to provide the actual working of the system Data Confidentiality: Not mentioned Access Control: Not mentioned
BaDS: Blockchain-Based Architecture for Data Sharing with ABS and CP-ABE in IoT [7]	IoT data	Hyperledger fabric and cloud storage PBFT is used for consensus	Access control: Smart contract for generating permission access control table Data confidentiality: ABS and CP ABE is used for signature and encryption User anonymity: Not required as it is a permissioned Blockchain
Ushare: user controlled social media based on blockchain [34]	Social media data	Ethereum	Access control: Smart contract for access control that performs transactions in the form of record of ownership of data items and number of shares that can be made on blockchain. Uses Distributed hash table

(*continued*)

Table 2. (*continued*)

Paper title	Use case	Blockchain framework used	Privacy Parameters
A Personal Privacy Data Protection Scheme for Encryption and Revocation of High-Dimensional Attribute Domains [9]	Data privacy	Hyperledger Fabric	Data Confidentiality: High-dimensional Attribute Domain-based Access Control: Attribute Revoking mechanism using sentry mode
Data privacy protection model based on blockchain in mobile edge computing [5]	Mobile edge computing		User Anonymity: Not addressed Data Confidentiality: Not addressed Access Control: Ciphertext policy attribute based encryption through RSA encryption
A Study of the Applicability of Ideal Lattice-Based FullyHomomorphic Encryption Scheme to Ethereum Blockchain [30]	Homomorphic encryption	For data confidentiality	Requirement of Third party for storing keys and use of off chain for managing the encryption
A Data Sharing Scheme Based on Blockchain System and Attribute-Based Encryption [7]	Attribute Based Encryption	data encryption and fine-grained access control	Researchers have used CP-ABE encryption for fine-grained access control using smart contract as well as data encryption

As seen from the Table 2, most of the privacy preservation techniques applied are on top of private blockchain and can cover up to only one or two data privacy factors. Non-financial applications named as Dapps are based on Public blockchain Ethereum and all three privacy factors implementation as in Fig. 1 need to be integrated to provide a secured layer above the Blockchain layer.

3 Proposed Methodology

Even though blockchain is designed to be immutable, there is security vulnerabilities associated with blockchain networks. A vulnerable program, for instance, can continue to give unwanted access to the ledger through the network or directly to its disk. As a

Table 3. Privacy Preservation Parameters Analysis

Papers	Confidentiality	Anonymity	Access Control
[4]	Available	Not available	Available
[2]	Available	Not available	Available
[3]	Available	Not available	Available
[35]	Not available	Available	Not available
[15]	Not available	Not available	Available
[36]	Available	Not available	Available
[34]	Not available	Not available	Available
Proposed System	Available	Available	Available

result, businesses must control, monitor, and prohibit unwanted access to their networks. Access could also be blocked by compromised, misplaced, or stolen encryption keys. The identity of blockchain nodes and participants must follow verification and authentication to prevent threat actors from imposing as valid users. The existing work is lacking in implementing the privacy parameters in combination and integration. From research it seems that blockchain can be used at its full potential only if the combination of the privacy parameters namely user anonymity, data confidentiality and data access are achieved by implementing their respective mechanism We proposed the use of Ring signature for user anonymity, homomorphic encryption for data confidentiality and smart contract-based access control for data access.

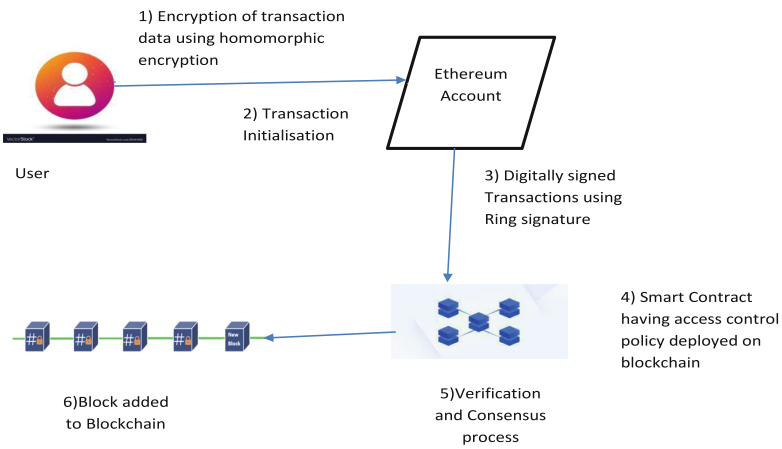


Fig. 3. Privacy Preservation at Architectural level

As shown in Fig. 3, the proposed system will consists of the following steps.

- Encryption: The user will first encrypt the data using homomorphic encryption

- Initialization: Transaction initialization through Ethereum account
- Anonymization: The transaction will be signed using a ring signature
- Access Control: Role based Smart contract for fine-grained access control such create, read, and update. Validation of the transactions as per the access control list and addition into the current block.
- Consensus: Verification and consensus of the block on Ethereum blockchain network.
- After successful execution of step5, block gets added into the blockchain.

3.1 Cryptographic Approaches for Privacy

Encryption Mechanism: Homomorphic encryption is used as an encryption mechanism in the proposed system. It is a method of encryption that allows processing on encrypted data. Any functions can be executed on encrypted data without the need of decryption. It uses a public key cryptographic algorithm such as RSA or ECC to encrypt and decrypt data. Only the authorized person with the valid private key can decrypt the data whereas an unauthorized person who does not have a private key can perform a variety of computations (or operations) algebraically on the encrypted data without actually decrypting the data. This property of homomorphic encryption sets it apart from other forms of encryption such as symmetric encryption and asymmetric encryption. Authorized parties can process and analyze the encrypted data without revealing its actual contents, thus protecting data privacy. Encrypted data remains secure, and the integrity of the process is maintained. This is important as the transaction data needs to be verified and homomorphic encryption can help in verification for blockchain transactions without decryption that is not possible with other encryption techniques.

Anonymity method: Out of the anonymity methods in the systematic review ring signature is used in the proposed system. This is because it provides a cryptographic way of handling anonymity making it applicable in the case of a Non-Financial blockchain based application. Basically in ring signature a group of people make a ring by creating a set of their public keys and the signature is generated with the help of private key of signer and this set of public keys. The verifier is able to verify the signature without knowing exactly who has signed the message. Ring signature makes it virtually impossible to identify the sender and recipient in a transaction. It can protect the identity of transaction initiators and the signatory's identity remains hidden during data exchange. The use of ring signature will ensure transaction unlinkability and provide complete authentication as well as non-repudiation as expected for any transaction in a blockchain environment. This type of anonymity is essential to overcome the problems encountered due to pseudo-anonymity used in blockchain network.

Access Control Method: Smart contract policies based on the Ciphertext based attribute encryption method can be used for granular level access control. Smart contracts can be written to include the access privileges as per role based access control list, and accordingly the transactions will be validated by the smart contract deployed on the blockchain network. This will provide fine-grained access privileges to the users of the system. Smart contracts define access rules and data sharing policies, Due to its Turing complete property; smart contracts enforce access control and define data-sharing policies. Fine-grained permission settings ensure that only authorized entities can access and utilize the encrypted data.

As shown in Fig. 3, integrating and combining above mentioned techniques, a separate protection layer can be added on the top of secured and immutable blockchain architecture to enhance privacy preservation in Decentralized applications. As per the current state of art, the proposed systems uniquely cover all three privacy preservation parameters using cryptographic and standard approaches to provide enhanced privacy for Non-Financial blockchain applications.

Potential weaknesses and challenges: Each of the cryptographic approaches need further analysis with respect to different symmetric cryptographic algorithms such as AES, DES and asymmetric cryptographic algorithms such as RSA, ECC to provide efficient and robust solution. Management of keys used in different algorithms is an important consideration while selecting the right one for the proposed system. As the system grows in size, scalability issue may arise which can be solved using IPFS as well as off blockchain databases. Smart contract are software code written using programming language i.e. solidity. They are subjected to security vulnerabilities; hence it is necessary to formally validate the smart contracts before using them in the proposed system and test them for possible security risks. The proposed system is based on Ethereum network that uses proof of stake consensus for block verification. Proof of stake helps to reduce energy consumption as it does not needs to solve complex puzzle as compared to mining in Bitcoin based application. Due to addition of privacy layer, other potential challenges such as computational overhead, space and time complexity are another important concern that requires careful inspection and analysis in the development of the proposed system. Authors are interested in continuing this research to address the above mentioned challenges and provide critical analysis as future work.

As shown in Table 3, the existing blockchain applications are providing maximum two privacy preservation features and the proposed methodology aims to include all three parameters to facilitate enhanced privacy to DApps.

4 Conclusion

Transaction immutability, audit trails and traceability, and consensus methods for transaction commit are some of the features of blockchain technology. Decentralized and non-financial applications using Ethereum blockchain are picking up speed due to the integral benefits of blockchain technology. Data privacy remains an important aspect in such applications. Blockchain in its inherent applicability does not provide privacy preservation. Privacy preservation can be achieved by implementing user anonymity, data confidentiality and access control. Different privacy methods are being used by Dapps developers, but the techniques are implemented in isolation or in combination of atmost two of these techniques. The authors proposed integration of three different methods for the three different privacy parameters to achieve a more robust, secure and privacy enhanced solution.

References

1. Singh, R., Dwivedi, A.D., Srivastava, G., Chatterjee, P., Lin, J.C.-W.: A privacy-preserving internet of things smart healthcare financial system. IEEE Internet Things J. **10**(21), 18452–18460 (2023)

2. Tomaz, E.B., Nascimento, J.C.D., Hafid, A.S., De Souza, J.N.: Preserving privacy in mobile health systems using non-interactive zero-knowledge proof and blockchain. IEEE Access **8**, 204441–204458 (2020)
3. Dagher, G.G., Mohler, J., Milojkovic, M., Marella, P.B.: Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. Sustain. Cities Soc. **39**, 283–297 (2018)
4. Puneeth, R.P., Parthasarathy, G.: Blockchain-based framework for privacy preservation and securing EHR with patient-centric access control. Acta Informatica Pragensia (2024). https://doi.org/10.18267/j.aip.225
5. Wu, J., Bu, X., Li, G., Tian, G.: Data privacy protection model based on blockchain in mobile edge computing. J. Softw. Pract. Exp. (2024)
6. Zhang, R., Xue, R., Liu, L.: Security and privacy on blockchain. ACM Comput. Surv. **52**(3), 34 (2020). Article 51
7. Zhang, Y., He, D., Choo, K.-K.R.: BaDS: blockchain-based architecture for data sharing with ABS and CP-ABE in IoT. Wirel. Commun. Mob. Comput. **2018**, 9 (2018). Article ID 2783658
8. Khan, Y., Varma, S.: An intelligent blockchain and software-defined networking-based evidence collection architecture for cloud environment. Sci. Programm., 1–19 (2021). https://doi.org/10.1155/2021/7294206
9. Wang, C., Lu, J., Li, X., Cao, P., Zhou, Z., Wen, Q.: A personal privacy data protection scheme for encryption and revocation of high-dimensional attribute domains. IEEE Access **11**, 82989–83003 (2023)
10. Xiaofeng, L.U., Songbing, F.U.: A trusted data access control scheme combining attribute-based encryption and blockchain. Netinfo Secur. **21**(3), 7–8 (2021)
11. Tariq, M.U.: Revolutionizing health data management with blockchain technology: enhancing security and efficiency in a digital era (2024). https://doi.org/10.4018/979-8-3693-1214-8.ch008
12. Gordon, W.J., Catalini, C.: Blockchain technology for healthcare: facilitating the transition to patient-driven interoperability. Comput. Struct. Biotechnol. J. **16**, 224–230 (2018)
13. Voloch, N., Gal-Oz, N., Gudes, E.: A trust based privacy providing model for online social networks. Online Soc. Netw. Media **24**, 100138 (2021). ISSN 2468-6964
14. Murimi, R.M.: A blockchain enhanced framework for social networking. Ledger (2019). https://doi.org/10.5195/ledger.2019.178
15. Zhang, S., Yao, T., Sandor, V.K.A., Weng, T.-H., Liang, W., Su, J.: A novel blockchain-based privacy-preserving framework for online social networks. Connect. Sci. **33**(3), 555–575 (2021)
16. Frimpong, S.A., et al.: RecGuard: an efficient privacy preservation blockchain-based system for online social network users. Blockchain Res. Appl. **4**(1), 100111 (2022)
17. Chen, N., Cho, D.S.-Y.: A blockchain based autonomous decentralized online social network. In: IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), Guangzhou, China, pp. 186–190 (2021)
18. https://diacc.ca/wp-content/uploads/2020/03/DIACC-White-Paper_Consumer-Digital-Identity-Leveraging-Blockchain_Feb2020.pdf
19. Zyskind, G., Nathan, O., Pentland, A.: Decentralizing privacy: using blockchain to protect personal data. In: IEEE Security and Privacy Workshops, pp. 180–184 (2015)
20. Salman, T., Zolanvari, M., Erbad, A., Jain, R., Samaka, M.: Security services using blockchains: a state of the art survey. IEEE Commun. Surv. Tutorials **21**(1), 858–880 (2019)
21. Godyn, M., Kedziora, M., Ren, Y., et al.: Analysis of solutions for a blockchain compliance with GDPR. Sci. Rep. **12**, 15021 (2022)
22. Belotti, M., Božić, N., Pujolle, G., Secci, S.: A Vademecum on blockchain technologies: when, that, and how. IEEE Commun. Surv. Tutorials **21**(4), 3796–3838 (2019)

23. Conti, M., Sandeep Kumar, E., Lal, C., Ruj, S.: A survey on security and privacy issues of bitcoin. IEEE Commun. Surv. Tutorials **20**(4), 3416–3452 (2018)
24. Banupriya, S., Kottilingam, K.: An analysis of privacy issues and solutions in public blockchain (Bitcoin). In: 2021 2nd International Conference for Emerging Technology (INCET), Belagavi, India, pp. 1–7 (2021)
25. Benhamouda, F., et al.: Can a public blockchain keep a secret? In: Pass, R., Pietrzak, K. (eds.) TCC 2020. LNCS, vol. 12550, pp. 260–290 (2020). https://doi.org/10.1007/978-3-030-64375-1_10
26. Kumar, E.S.: Preserving privacy in Ethereum blockchain. Ann. Data Sci. **9**(4), 675–693 (2020)
27. https://ethresear.ch/t/the-application-of-zk-snarks-in-solidity-privacytransformation-computational-optimization-and-mev-resistance/17017
28. Mahmood, Z., Vacius, J.: Privacy-preserving block-chain framework based on ring signatures (RSs) and zero-knowledge proofs (ZKPs). In: International Conference on Innovation and Intelligence for Informatics, Computing and Technologies (3ICT), Sakheer, Bahrain, pp. 1–6 (2020)
29. Wang, L., Peng, C., Tan, W.: Secure ring signature scheme for privacy-preserving blockchain. Entropy **25**(9), 1334 (2023)
30. Chen, P.-C., Kuo, T.-H., Wu, J.-L.: A study of the applicability of ideal lattice-based fully homomorphic encryption scheme to ethereum blockchain. IEEE Syst. J. **15**(2), 1528–1539 (2021)
31. Liu, G., Xie, H., Wang, W., Huang, H.: A secure and efficient electronic medical record data sharing scheme based on blockchain and proxy re-encryption. J. Cloud Comput. **13**(1) (2024)
32. https://tornado.ws/
33. Bashi, I.: Mastering Blockchain. 2nd edn. Packt Publishing (2018)
34. Chakravorty, A., Rong, C.: Ushare. Association for Computing Machinery, January 2017. Art. no. 99
35. Yarava, A., Bindu, C.S.: Ring signature-based blockchain for guaranteeing privacy preservation in online social networks. Multimedia Tools Appl. **83**(16), 47151–47174 (2023)
36. Jiang, L., Zhang, X.: BCOSN: a blockchain-Based decentralized online social network. IEEE Trans. Comput. Soc. Syst. **6**(6), 1454–1466 (2019)

Identifying Application-Level Distributed Denial-of-Service (DDoS) Attacks with Machine Learning (ML)

Muhammad Atif Yaqub[1], Muzzamil Mustafa[2], Saad Hussain Chuhan[1], Basit Sattar[2], Muhammad Zulkifl Hasan[3(✉)], Muhammad Zunnurain Hussain[4], Muhammad Saad Qaiser[3], Zain ul Abdin[3], Afshan Bilal[1], and Jibran Ali[5]

[1] Department of Computer Science, National College of Business Administration and Economics, Lahore, Pakistan
[2] Department of Artificial Intelligence, University of Management and Technology, Lahore, Pakistan
[3] Faculty of IT and Computer Science, University of Central Punjab, Lahore, Pakistan
Zulkifl.hasan@ucp.edu.pk
[4] Department of Computer Science, Bahria University Lahore Campus, Lahore, Pakistan
[5] Premier Systems, Karachi, Pakistan

Abstract. The modern digital age has seen an exponential growth in internet penetration, which in turn has increased the probability of a denial-of-service attack. When an evil user launches a Denial of Service (DoS) attack, it's with the goal of flooding computing and network resources and blocking access to them for normal users. These kind of assaults may happen at any level of the OSI model, from the application layer all the way down to the transport layer. This research focuses on application layer DoS attacks instead of transport and network DoS attacks and attempts to detect Denial of Service using Machine Learning and Neural Network approaches. An updated dataset of Denial of Service attacks divided into many subcategories was used in this experiment. According to the research, the decision tree approach performed well in terms of accuracy and splitting compared to the logistic regression approach.

Keywords: Denial of Service (DoS) · Denial of Service (DDoS) · Machine learning (ML) · Neural Network (NN)

1 Introduction

Networking has changed a lot in the last ten years partly because of the exponential increase in the risk that is introduced to the network environment as new technologies are developed. One way that original users are prevented from accessing a system or network is through Denial-of-Service by malicious actors. DoS attacks are attacks that try to send as much traffic as a network can handle so that they cannot handle original traffic. The description of denial-of-service attacks has led to the development of ways to implement and prevent it. New advancements that use machine learning have developed

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 105–115, 2025.
https://doi.org/10.1007/978-3-031-86305-9_9

ways to mitigate and detect the occurrence of DoS. Machine learning methods allow for the power for which machine learning algorithms to use data mining in order to detect the patterns that show a given form of attack. A DoS attack may be found and prevented using this description by focusing on the three types of origin, instinct, and duration. DoS incidents may be found and prevented as a result of the development that machine learning has achieved. Denial of Service acts is the first and most serious hazard to the network. DoS attackers leak data from the network layer to the application layer. The level of use, the complexity with which they are made, and the size of the assaults are increasing, according to analysis. A DoS attack is when the attacked system becomes idle and loses resources on a targeted machine. Thus, it is necessary to provide IDS that will protect sensitive information and deal with the DoS in a rather collected way. The most common DoS attacks are ICMP flooding, UDP flooding, SYN flooding, and HTTP flooding. One of the examples is that of the DoS attack of UDP flooding. The overflow of so many packets is being sent to a specific target. It provides an overwhelming amount of attacks with random data that are sent by the attacker. Another example is the ICMP flooding that is sending most of the ICMP Echo replies to slow down the system. The slowing does not provide enough time for the target to respond. When SYN flooding is implemented correctly, a target may be overwhelmed by the flood to the point where it cannot respond to regular network traffic. The final one is the HTTP flooding, which receives an overwhelming amount of requests that make is able to drop off the real requests. The identification of Denial of Service (DoS) attacks has been a hotspot for research because of the damage they may do. They have been identified using a number of different detection approaches, some of which use machine learning algorithms like logistic regression and decision trees. Some interesting results have emerged from an examination of a dataset including DoS/DDoS assaults.

2 Literature Review

Datasets created for the express goal of researching and evaluating application layer DoS attacks are plentiful when one looks into datasets concentrating on such attacks. A variety of datasets are available, covering a broad range of topics, such as application layer DoS attack activity and network traffic.

The most comprehensive dataset available at the moment is the KDD Cup 99 dataset. This encompasses both benign network activity and malicious traffic resulting from Denial-of-Service (DoS) attacks. Pod, smurf, teardrop, back, land, Neptune, and forty-one more sorts of attacks are included in the 4898 connection records that make up the collection. If you're inclined to study application layer attack threats on different DoS forms, I could recommend the following datasets. The NSL-KDD dataset, which includes 125,979 connection records from which we are certain that 22 different attack techniques were utilized. However, there are no instances of ordinary traffic in this dataset; only the traffic from DoS attacks. In studying application layer DoS attacks, SQL injection, and other forms of cross-site scripting, the ADFA-LD dataset is a unique phenomenon. This much assault occurrences are present in the collection, which corresponds to 13 different kinds of attacks. The DDoS-Attack 2018 dataset is perhaps the best to use for research concerning application layer DoS attacks. The collection encompasses new

data on additional application layer assaults such as web service assaults and distributed denial of service. In all, there are 2,540,316 samples from a total of 35 distinct kinds of assault in the collection. With regard to the DDoS-Attack-2018 dataset, other forms of application layer assault new measurements forms like SYN flood, HTTP flood, and UDP flood are included. The dataset includes 6,974,337 samples from a total number of 12 different kinds of attacks.

The following are some of the Literature Reviews Regarding the Application Layer DoS Attack Dataset. The "A Novel Intrusion Detection System Based on Machine Learning for Denial of Service Attack Detection" paper was authored by Jiwei Liu and friends in 2017. An IDS based on machine learning which is capable of identifying denial-of-service assaults is proposed in the study. A classification method known as SVM is utilized by the intrusion detection system in order to identify the harmful one. Additionally, the research reports on the effectiveness of the intrusion detection system using a data set that includes denial of service attacks from the real world. "Dos Attack Detection Utilizing Neural Networks" was authored by H. P. and friends in 2002. A neural network-based approach for the detection of dos is presented in this paper. In this case, the usage of a neural network is the means to avert these types of attacks. A neural network with multiple layers of perceptron's is used to detect the assault that is directed towards dos. The study reveals that the neural network is capable of detecting dos attacks with an accuracy of over 99.9%. "Detecting DoS Attacks Using Machine Learning Techniques" was published by M. M. Hasan and friends in 2017. The main objective of the project is to produce a system based on machine learning that can be used to identify denial of service attacks. The creation and presentation of the model incorporating random forest and deep learning methods can put risk into perspective and assign it a number based on the likelihood of it causing a denial of service assault. The results of the study indicate that the proposed approach is capable of detecting assaults in which it has a high detection rate of over 97%. In summary, the literature review has managed to cover several datasets on application layer Dos attacks. From what I have observed, each of staging these datasets, I can say, they are a major achievement. KDD Cup 99's datasets for the are by far the largest that now exist. While the DDoS-Attack-2018 dataset is the most current of the existing large datasets. Since they all provide a new perspective to the analyst on how application layer Dos attack threats are, one can further carry out their research and analyze after gaining this information.

3 Application-Layer DDoS Dataset

The Application-Layer Distributed Denial of Service (DDoS) dataset is a dataset that contains various information from different sources and comprises data related to DDoS attacks at the application layer. The dataset comprises data that includes two types of application-layer DDoS attacks, namely SYN flood and HTTP flood. We also include a collection of Slowloris data here. Each entry contains information about the type of attack, source IP address, target IP address, target port, attack strength, attack duration, and attack timestamp. This dataset is of interest to researchers studying the characteristics and nature of application-layer DDoS attacks. It can be employed to evaluate the typical parameters of the attacks, patterns, and other important aspects of this DDoS type. The

dataset might also be used to create and verify novel types of defenses against application-layer DDoS attacks. By combining data collected from multiple sources, the dataset was created as a unique collection of approximately 200,000 entries from 2017 to 2019. This dataset is available to the public on the internet in a csv format. The Application-Layer DDoS dataset is an invaluable resource for researchers and professionals interested in studying application-layer attacks or developing their approaches to application-level DDoS mitigation.

Table 1 consists of several features that are present in the dataset.

Table 1. Attributes in Dataset

Attribute Name	Attribute Name
FlowID BwdPackets/s SourceIP	FlowIATStd BwdAvgPackets/Bulk
MinPacketLength SourcePort	FlowIATMax BwdAvgBulkRate FlowIATMin
MaxPacketLength DestinationIP	SubflowFwdPackets FwdIATTotal
PacketLengthMean DestinationPort	SubflowFwdBytes FwdIATMean
PacketLengthStd Protocol	SubflowBwdPackets FwdIATStd
PacketLengthVariance Timestamp	SubflowBwdBytes FwdIATMax
FINFlagCount FlowDuration SYNFlagCount	InitWinbytesforward FwdIATMin
TotalFwdPackets RSTFlagCount	InitWinbytesbackward BwdIATTotal
TotalBackwardPackets PSHFlagCount	actdatapktfwd BwdIATMean
TotalLengthofFwdPackets ACKFlagCount	minsegsizeforward BwdIATStd ActiveMean
TotalLengthofBwdPackets URGFlagCount	BwdIATMax ActiveStd BwdIATMin
FwdPacketLengthMax CWEFlagCount	ActiveMax FwdPSHFlags ActiveMin
FwdPacketLengthMin ECEFlagCount	BwdPSHFlags IdleMean
FwdPacketLengthMean Down/UpRatio	BwdPSHFlags IdleMean FwdURGFlags
FwdPacketLengthStd AveragePacketSize	IdleStd BwdURGFlags IdleMax
BwdPacketLengthMax AvgFwdSegmentSize	BwdHeaderLength IdleMin FwdPackets/s
BwdPacketLengthMin AvgBwdSegmentSize	Label
BwdPacketLengthMean FwdHeaderLength	actdatapktfwd BwdIATMean
BwdPacketLengthStd FwdAvgBytes/Bulk	minsegsizeforward BwdIATStd ActiveMean
FlowBytes/s FwdAvgPackets/Bulk	BwdIATMax ActiveStd BwdIATMin
FlowPackets/s FwdAvgBulkRate	ActiveMax FwdPSHFlags ActiveMin
FlowIATMean BwdAvgBytes/Bulk	InitWinbytesforward BwdIATTotal
FwdIATMin InitWinbytesbackward	actdatapktfwd BwdIATMean
BwdIATMax ActiveStd BwdIATMin	minsegsizeforward BwdIATStd ActiveMean
ActiveMax	FwdPSHFlags ActiveMin

This Fig. 1 shows the Training of the dataset with these several features.
This Fig. 2 shows the testing of the dataset with these several features.

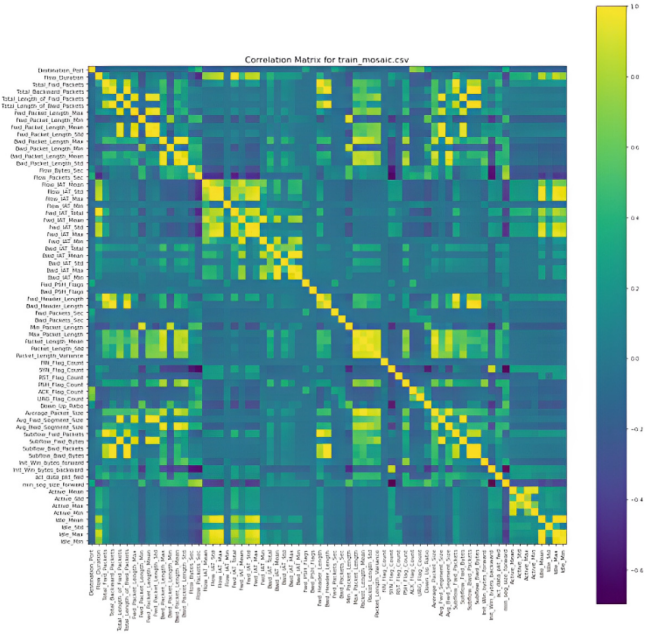


Fig. 1. Training Dataset

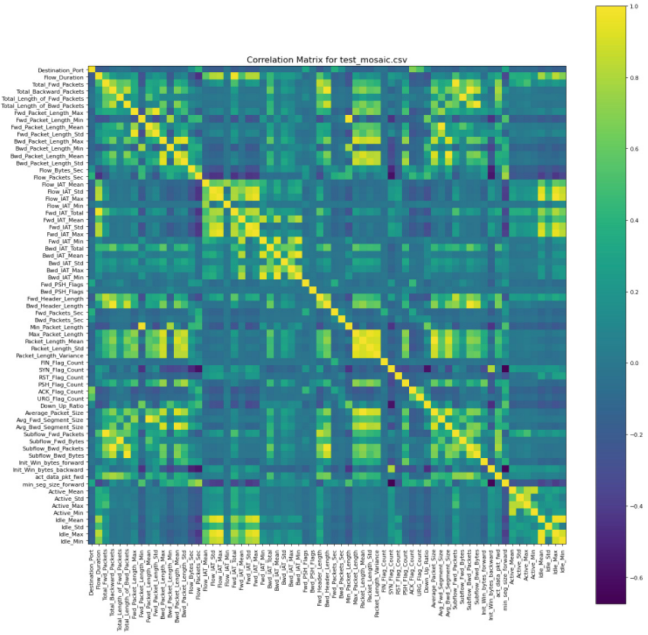


Fig. 2. Testing Dataset

4 Research Objectives and Outcomes

The goal of this study is to create an optimum Machine Learning method for analyzing Denial of Service (DoS) attacks using Decision Tree and Logistic Regression. We will use the DDoS Application Level Dataset to not only assess protection mechanisms against different DDoS tactics, but also to detect pattern features of DDoS assaults.

5 Research Contributions

This study provides the following significant contributions:

- By comparing the performance of machine learning methods, this research makes innovative contributions to the subject of cybersecurity.
- Detecting DoS assaults and underlining the Decision Tree algorithm's advantages in this context.
- These results have practical relevance for security experts and enterprises looking for effective anti-malware protection methods.

6 Experimentation

6.1 Proposed Methodology

This work developed from the DDoS-Attack-2018 dataset to identify the application layer DDoS attacks, by the help of Decision Tree and Logistic Regression models. Steps that were taken in procuring the dataset for analysis included the handling of missing values, categorizing of variables and numerical scaling. Four fold cross validation was used as a means of evaluation, whereby 80% of data was used for testing while 20% was used for training the model. In the process of hyperparameter tuning, grid search was used with such parameters: the depth of the Decision Tree and the regularization parameter for Logistic Regression. Cross validation five folds were used in training and evaluating both the models. Decision Tree aimed at classifying the traffic as either benign or malicious by learned attack patterns and Logistic Regression predicted the probability of an attack based on the input attributes. An evaluation of the developed models was established based on accuracy, precision, recall, and F1-score. We implemented the models using the Weka software and populated it with network traffic data; then compare and contrast the final results to see which model performed best in DDoS alerting.

6.2 Decision Tree

In the area of information security, decision trees help to identify dangers or threats. A decision tree diagram is a series of verdicts with each branch displaying one option. Using a decision tree, security experts may easily categorize threats and resolve them as needed. A decision tree may be used to uncover transgressions by recognizing abnormalities in network data flow. In conclusion, another application of a decision tree is for creating information rules that help safeguard sensitive information. However, the highest accuracy of the Decision tree algorithm is 99% is obtained in the model after testing (Fig. 3).

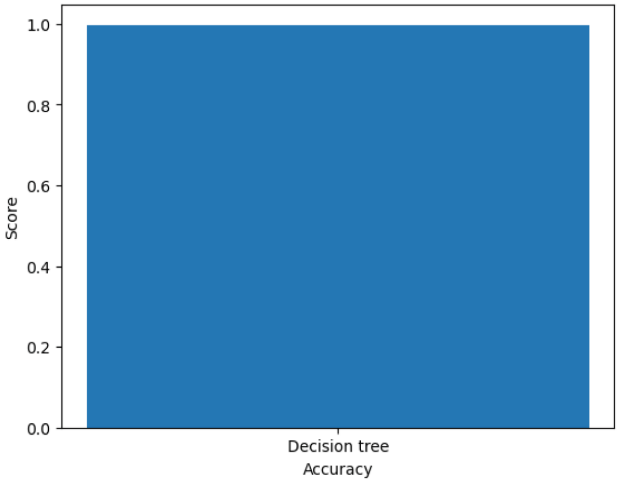


Fig. 3. Shows the accuracy of Decision Tree

6.3 Logistics Regression

Logistic regression can also be employed to detect an intrusion. Consider it a way to extrapolate the likelihood of an incursion focused on a set of input parameters. A model that applies logistic regression to a collection of traffic from an incursion and traffic from harmless connections may offer an accurate forecast of whether a particular link is infiltrated. Another application could be figuring out which attributes are most related to the probability of a successful incursion measure what to focus on a high-security plan. For these reasons above, our Logistics' Regression offers an accuracy of 95% which also falls within the acceptable range (Fig. 4 and Tables 2, 3).

7 Result Analysis and Discussion

Decision tree and Logistics Regression methods were employed in this work to identify Denial of Service (DoS) attacks. This graph depicts the Decision tree's accuracy after being trained on diverse behaviors using 80 attributes. The same dataset was exposed to Logistic Regression (i.e. benign or DoS packet), which was initially constituted of different types of DoS assaults and then turned into a binary form. The accuracy of the suggested approach was determined by measuring the ratio of true positives and true negatives to the total number of instances.

A True Positive is a record that has been appropriately categorized as a positive label record. A True Negative, on the other hand, is a record that has been appropriately classified as a negative label record.

The sum of True Positive and True Negative is defined as Condition Positive.

False Positive happens when a positively labeled record is incorrectly recognized as a negative record.

V = False When a negative label record is wrongly classed as a positive label record, the result is negative.

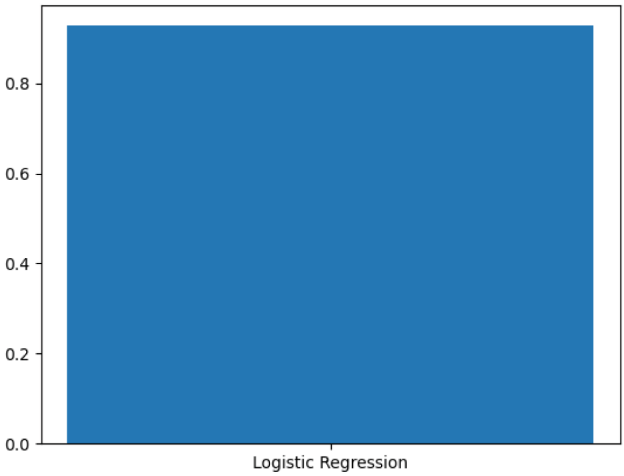


Fig. 4. Shows the accuracy of Decision Tree

Table 2. Accuracy of Logistics regression

Sr No	# Training Records	# Tested Records	Accuracy
1	3910(30%)	7744	95.8783%
2	6143(40%)	6797	95.5099%
3	8191(50%)	5818	95.8760%
4	10239(60%)	4853	95.8818%
5	12335(70%)	3884	95.8956%
6	14387(80%)	2915	95.8956%

Table 3. Accuracy of Decision Tree

Sr No	# Training Records	# Tested Records	Accuracy
1	3910(30%)	7744	99.9194%
2	6143(40%)	6797	99.9268%
3	8191(50%)	5818	99.9308%
4	10239(60%)	4853	99.9502%
5	12335(70%)	3884	99.9475%
6	14387(80%)	2915	99.9512%

7.1 Result Analysis for Decision Tree

We utilized 20% of the dataset for training at the start of the experiment. There was no significant change in detection rate between the 30% and 70% learning scores. Following testing, it was discovered that 30% of the dataset produced the greatest accuracy. It was also discovered that as the number of packets utilized for training increased, so did the accuracy.

The decision tree analysis for identifying DDoS assaults may be fairly sophisticated. The decision tree may be trained on a dataset of previously classified attack traffic and then used to classify fresh traffic.

7.2 Result Analysis for Logistics Regression

We ran an experiment with numerous records from the dataset, using 20–80% of the records for training.

This method is appropriate for datasets with a few independent variables and a dependent variable with just two potential outcomes, such as yes/no, true/false, and so on. The Logistic Regression model may be used to distinguish malicious activity from benign traffic in the context of a DDOS assault.

7.3 Discussion

The Application-Layer DDoS Dataset was chosen since it solely includes DoS assaults and regular packets. To identify the appropriate split of training and test data, we examined two techniques: decision tree and logistics regression methods. A larger number of learning packets resulted in a higher detection rate, according to our data. With 30% training records, the Decision tree algorithm produced a greater accuracy of 99.57%, whereas the Logistics regression method achieved a lesser accuracy of 95.18% with 50% training. The suggested system revealed that the Decision tree approach outperforms Logistics regression in identifying DoS assaults at the application layer. As a possible future work, considering the research direction of a multi-class classification to differentiate among different classes of DDoS attacks by applying the abovementioned ML algorithms at the application layer, transport layer and network layer. In multi-class there is other algorithms we can apply like Random Forest, Support Vector Machines, Neural Networks among others.

8 Conclusion and Future Works

This study was able to effectively explain and show the application of Machine Learning algorithms: Decision Tree and Logistic Regression in the identification of application layer DDoS attack. Indeed, Decision Tree which has been used in the analysis characterized better results than Logistic Regression model in the aspect of accuracy that confirms suitability of this approach for this type of threats identification. Nevertheless, the study targeted application-layer DDoS attack category only, without making distinction between Heartbleed, slowhttptest, slowloris, and HTTP flood types of the

attack. Furthermore, no solutions for transport and network layer attacks were found. Future work should diversify the two categories of the attacks and apply multi-class classification where relevant using the machine learning algorithms such as; Random forest & Neural Networks. In addition, feature reduction methods may be used to filter out unnecessary features such as PCA or RFE to reduce the number of the features in the set which may enhance model and increase its overall generality. Finally, it is crucial to assess the performance of the constructed system compared to the transport and network layer DDoS attack in a larger set in order to enhance the efficiency of protection means. This would improve the operational relevance of the approach more especially in operational cybersecurity settings.

References

1. Ahmed, M., Abdel-Ghaffar, W., Abdelhamid, M.M.: A literature review of application layer DoS attack datasets. In: 2018 8th International Conference on Cyber Security and Protection of Digital Services (Cyber Security), pp. 1–7. IEEE (2018)
2. Yao, Y., Ye, D., Liu, Y.: Application layer DoS attacks and defense strategies: a review of literature. IEEE Access **8**, 155361–155376 (2020)
3. Khan, M., Zafar, A., Chaudhary, A.: Application layer denial of service (DoS) attacks: a review. Int. J. Eng. Technol. **8**(2.3), 541–544 (2019)
4. Srinivas, M., Mallya, S.: Application layer denial of service attack detection and prevention techniques: a survey. Int. J. Comput. Trends Technol. **59**(2), 6–15 (2019)
5. Alshayeb, M., Al-Nabki, M.: Application layer denial of service attacks: a comprehensive survey. Procedia Comput. Sci. **94**, 1178–1187 (2016)
6. Alghamdi, M., Al-Shayeb, M.: A survey of application layer denial of service attacks and countermeasures. Int. J. Comput. Netw. Commun. (IJCNC) **7**(2), 30–45 (2015)
7. Tsang, G.C., Chan, P.P., Yeung, D.S., Tsang, E.C.C.: Denial of service detection by support vector machines and radial-basis function neural network. In: Proceedings of 2004 International Conference on Machine Learning and Cybernetics, vol. 7, pp. 4263–4268. IEEE (2004)
8. Seufert, S., O'Brien, D.: Machine learning for automatic defence against distributed denial of service attacks. In: IEEE International Conference on Communications, ICC 2007, pp. 1217–1222. IEEE (2007)
9. Umarani, S., Sharmila, D.: Predicting application layer DDoS attacks using machine learning algorithms. Int. J. Comput. Control Quantum Inf. Eng. **8**(10) (2014)
10. Subbulakshmi, T., BalaKrishnan, K., Mercy Shalinie, S., AnandKumar, D., Ganapathi Subramanian, V., Kannathal, K.: Detection of DDoS attacks using enhanced support vector machines with real time generated dataset. In: 2011 Third International Conference on Advanced Computing (ICoAC), pp. 17–22. IEEE (2011)
11. Tan, Z., Jamdagni, A., He, X., Nanda, P., Liu, R.P., Hu, J.: Detection of denial-of-service attacks based on computer vision techniques. IEEE Trans. Comput. **64**(9), 2519–2533 (2015)
12. Nosrati, E., Kashi, A.S., Darabian, Y., Tonekaboni, S.N.H.: Register flooding attacks detection in IP multimedia subsystems by using adaptive z-score CUSUM algorithm. In: 2011 International Conference on Information Technology and Multimedia (ICIM), pp. 1–4. IEEE (2011)
13. Bao, C.M.: Intrusion detection based on one-class SVM and SNMP MIB data. In: Fifth International Conference on Information Assurance and Security, IAS 2009, vol. 2, pp. 346–349. IEEE (2009)

14. Alkasassbeh, M., Al-Naymat, G., Hassanat, A.B., Almseidin, M.: Detecting distributed denial of service attacks using data mining techniques. Int. J. Adv. Comput. Sci. Appl. **7**(1) (2016)
15. Apale, S., Kamble, R., Ghodekar, M., Nemade, H., Waghmode, R.: Defense mechanism for DDoS attack through machine learning. Int. J. Res. Eng. Technol. **3**(10), 291–294 (2014)
16. Araar, A., Bouslama, R.: A comparative study of classification models for detection in IP networks intrusions. J. Theor. Appl. Inf. Technol. **64**(1) (2014)
17. Hasan, M.A.M., Nasser, M., Pal, B., Ahmad, S.: Support vector machine and random forest modeling for intrusion detection system (IDS). J. Intell. Learn. Syst. Appl. **6**(01) (2014)
18. Powers, D.M.: Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation (2011)
19. Khandelwal, M., Gupta, D.K., Bhale, P.: DoS attack detection technique using back propagation neural network. In: 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp. 1064–1068. IEEE (2016)
20. Kshirsagar, D., Sawant, S., Wadje, R., Gayal, P.: Distributed intrusion detection system for TCP flood attack. In: Singh, R., Choudhury, S. (eds.) Proceeding of International Conference on Intelligent Communication, Control and Devices, vol. 479, pp. 951–958. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-1708-7_111
21. Sharafaldin, I., Lashkari, A.H., Ghorbani, A.A.: Toward generating a new intrusion detection dataset and intrusion traffic characterization. In: Information Systems Security and Privacy, ICISSP (2018)
22. Kshirsagar, D.D., Sale, S.S., Tagad, D.K., Khandagale, G.: Network intrusion detection based on attack pattern. In: 2011 3rd International Conference on Electronics Computer Technology (ICECT), vol. 5, pp. 283–286. IEEE (2011)
23. Kshirsagar, D., Sawant, S., Rathod, A., Wathore, S.: CPU load analysis minimization for TCP SYN flood detection. Procedia Comput. Sci. **85**, 626–633 (2016)
24. Principe, J.C., Euliano, N.R., Lefebvre, W.C.: Neural and Adaptive Systems: Fundamentals Through Simulations, vol. 672. Wiley, New York (2000)
25. Distributed Denial of Service Trends Report (2017). https://www.verisign.com/enIN/securityservices/ddosprotection/ddosreport/index.xhtml. Accessed 01 Jan 2018

"Cybersecurity Threats and Trust Dynamics in Digital Payment Systems: An Analysis of Domestic Fraud and User Perspectives"

Vishal Ramaiya[1(✉)], Pramod Goyal[1], and Neeraj Kumar Dubey[2]

[1] Faculty of Management Studies, Marwadi University Rajkot, Rajkot, India
vishal.ramaiya119356@marwadiuniversity.ac.in
[2] ABS, Amity University MP, Gwalior, India

Abstract. The widespread adoption of electronic payment systems has introduced new cybersecurity challenges alongside its convenience. This paper examines key cybersecurity issues in digital payment platforms, particularly domestic fraud and the dynamics of user trust. The objective is to analyze how various cybersecurity threats, such as phishing, malware, and data breaches, impact the security and reliability of digital payments. The research specifically explores the influence of security breaches on user trust, while considering demographic factors such as age and education levels. In contrast to previous studies, this work provides a detailed analysis of how security breaches affect user confidence across different demographic groups. The study emphasizes the need for robust cybersecurity measures and improved security practices by financial institutions to foster trust and encourage broader adoption of digital payment systems. This research offers valuable insights into how demographic factors shape user perceptions of security, providing a foundation for more tailored cybersecurity strategies in the digital payments sector.

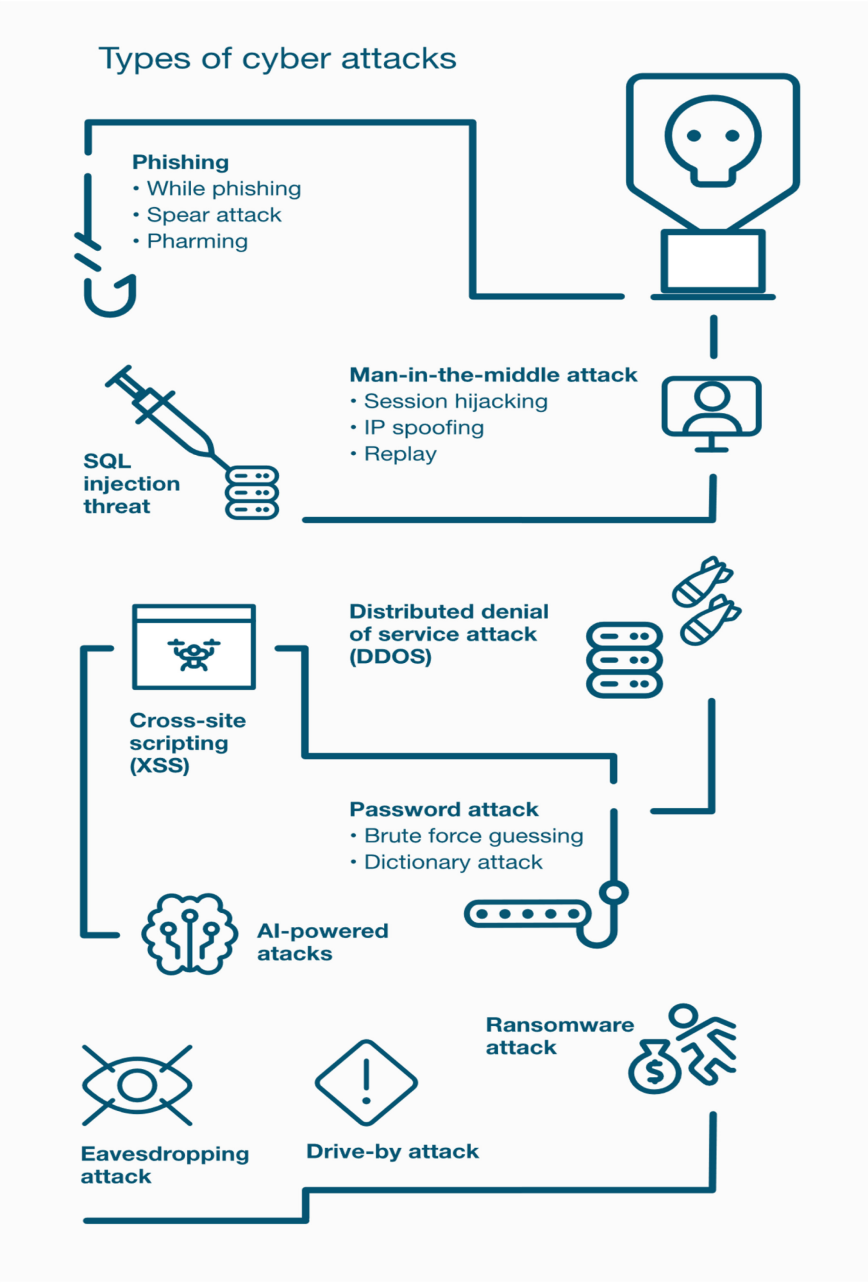
Keywords: Cybersecurity · Cyber Threats · FinTech Security · Domestic Payment Frauds · Trust Dynamics in Digital Payment

1 Introduction

Digital technology has revolutionized financial transactions, rendering digital payment systems a crucial component of modern economies. At a time when there is a general trend toward less and less use of cash in various transactions, mobile wallets and other forms of electronic payment systems have become very popular because of their simplicity and user-friendly interface. At the same time, however, this rapid shift presents more vulnerabilities, particularly in cybersecurity. In fact, the very fundamental benefits that make digital payments so appealing instantaneity, online connectivity, and widespread availability are the very factors that make them vulnerable to cyber threats.

V. Ramaiya—Research Scholar.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 116–138, 2025.
https://doi.org/10.1007/978-3-031-86305-9_10



(Source: yubico, 2024)

Major risks such as Internet fraud, identity theft, phishing, and malware attacks constitute significant barriers to digital payment systems, impacting their security and reliability. As the digital payment systems continue to grow in their complexity and breadth, so too are fraudsters developing new ways to exploit any weakness in them. More importantly than financial loss, a breach of this type can impact confidence in the service's use and hinder the development of digital financial services.

Furthermore, to ensure their effectiveness, payment systems in any country must adhere to rigorous standards of security, safety, dependability, effectiveness, and accessibility. The adoption of electronic payment systems, which facilitate online transactions for goods and services, varies across countries, each advancing at its own pace in response to the evolving landscape of digital commerce. Ensuring that these systems are secure and trustworthy is crucial for maintaining user confidence and supporting the broader adoption of digital payment solutions (Vishal Ramaiya, 2024).

Types of Cyber Security Threats

The expansion of digital payment systems has introduced various cyber security threats that jeopardize the security and reliability of financial transactions. Understanding these threats is vital for implementing effective protective measures. The following are some of the most significant types of cyber security threats:

1. **Malicious Software (Malware)**: malicious software that disrupts, destroys, or causes unauthorized access to computer systems. Viruses, worms, ransomware, and spyware all fall into the category of malware that can compromise sensitive data or badly disrupt system functions.
2. **Deceptive Phishing**: It is a fabricated scheme devised by hackers to dupe people for personal or proprietary information. The phishing, in general, entails sending fictitious emails, texts, or websites similar to the real ones with the intentions of drawing out information regarding password and credit card numbers among others.
3. **Service Disruption (DoS/DDoS Attacks)**: These are the attacks intended to swamp the system, service, or network with more than the normal quantity of traffic, hence making it unreachable. Such disruptions may pose a great deal of delay in operation, financial loss, or even reputational damage to a company.
4. **Interception Attacks (MitM)**: MitM occurs when a cybercriminal secretly intercepts and is even able to modify the communication between two entities. Such an attack could compromise data transfers and hence result in unauthorized access to it or some fraudulent action.
5. **Database Exploitation (SQL Injection)**: SQL injection is a practice where the hacker actually tries to inject malicious SQL code into the system's database in order to find loopholes that, when compromised, may lead to unauthorized access, tampering with data, or loss of the database altogether.
6. **Data Ransom (Ransomware)**: Ransomware is a sort of malware that securely locks or encrypts a user's data and renders them unreachable until the owner pays a ransom to the criminals. This can result in organizational monetary loss, loss of crucial data, and operational halt for failing to respond to the ransom.

7. **Unpatched Vulnerabilities (Zero-Day Exploits)**: Zero-day exploits are a type of attack that leverages software and hardware flaws that have previously been uncovered, during which time developers can release a patch. These are extremely dangerous since they could be used for unauthorized access or much worse before a patch is available.
8. **Persistent Intrusions (APTs)**: Advanced, long-term intrusions where the intruder successfully manipulates entry and ascertains access to a network. These attacks are done to steal sensitive data for a really long period of time and include high expertise and resources.
9. **Internal Risks (Insider Threats)**: Insider threats emanate from individuals internally in an organization who have legitimate access to sensitive systems and data. The threats can be deliberate threatening data, such as information theft or accidental, such as unintentional disclosure of data, but each has the potential for serious damages.
10. **Psychological Manipulation (Social Engineering)**: This involves the deliberate deception of people to break security through psychological manipulation. Impersonation, baiting, and pretexting are one of several methods used to gain unauthorized access or information.

By efficiently reducing these cyber security risks, enterprises can enhance the safeguarding of their digital payment systems and assure the security and reliability of financial transactions.

2 Significance of the Study

There are several dimensions to the importance of the research on "Cyber Security Dynamics in Digital Payment Systems: Insights from an Analytical Study." This initiative aims to increase knowledge about the many cyber security risks encountered by digital payment systems, thereby enlightening stakeholders about the need for implementing strong security protocols to safeguard sensitive financial data. The findings can provide valuable insights for politicians and regulatory authorities, guiding the formulation of more efficient policies that improve transaction security. By rectifying weaknesses, the research seeks to cultivate more user confidence, which is essential for the expansion of digital payment systems. Moreover, the suggested countermeasures function as important assets for financial organizations, enabling them to execute efficient security protocols. Furthermore, this research contributes to the scholarly discourse by laying the groundwork for future investigations into developing cyber security issues. Overall, the research underscores the need for cooperation among stakeholders to jointly tackle cyber security concerns, thereby improving the overall security stance of the digital payment ecosystem.

3 Research Gap

Despite the large amount of work on cybersecurity threats and digital payments, no study focused on how these threats influence the level of user trust considering factors like age, education, and occupation. Most of the studies are devoted to broad cybersecurity risks and their solutions but do not touch upon any issues related to the psychological and behavioral aspects of trust, especially during or after a security breach. Besides, the demographic characteristics were not comprehensively studied in relation to whether breach occurrences could happen. This paper seeks to fill in this gap by considering how cyber security breaches actually affect the dynamics of trust based on different demographics, thereby furthering the understanding on how to tailor security measures in order to enhance user trust and satisfaction in digital payment systems.

4 Objectives

1. To evaluate primary cyber security threats associated with digital payment systems.
2. To analyze the volume, value, and fraud rates of domestic payment frauds
3. To evaluate the influence of security breaches on the trust of users in digital payment systems.
4. To analyze users' satisfaction with digital payment systems in relation to perceived security and trustworthiness.
5. To examine the relationship between demographic factors and the likelihood of experiencing a security breach or unauthorized transaction on a digital payment platform
6. To propose strategies for enhancing the security of digital payment systems.

5 Hypothesis

1. Ho1: There is no significant relationship between experiencing a security breach and the trust in digital payment systems.
2. H02: There is no significant correlation between user satisfaction with digital payment systems and the perceived security and trustworthiness of these systems.
3. Ho3: There is no significant correlation between demographic factors and experiencing a security breach or unauthorized transaction through a digital payment platform.

6 Literature Review

Author(s), Year	Title of Study	Objective	Methodology	Key Findings
Sameer Patil, Sagnik Chakraborty (2019)	A Cybersecurity Agenda for India's Digital Payment Systems	To examine the growing cyber risks in digital payments	Qualitative analysis of cybersecurity incidents and regulatory frameworks	Identified key vulnerabilities and provided recommendations for government, business, and global measures
Arun Prasad G.S. & Dr. Arivazhagan R. (2021)	A Study on User Perceived Trust in Mobile Wallet	To investigate the users' perception towards Mobile Wallet's security and trust	Descriptive research design, online surveys, email surveys, and personal interviews	Identified five significant variables that influence trust among users, with technology playing a significant role
Moon et al. (2022)	Towards the Advancement of Cashless Transaction: A Security Analysis of Electronic Payment Systems	To analyze cybersecurity issues in electronic payment systems and establish methodologies for measuring security	Literature review of related studies on cybersecurity in electronic transactions	Identifies vulnerabilities in mobile payment apps and proposes methodologies for measuring security strengths
Dr. Bhupali Shah (2022)	A Review of Cyber Security Issues in Online Banking and Online Transactions	To review recent trends in cybersecurity issues related to online banking and transactions	Review-based methodology focusing on analyzing recent trends in cybersecurity issues	Online banking is a prime target for cybercriminals. Key precautions include checking for HTTPS, enabling two-factor authentication, and avoiding saving sensitive information on public computers

(*continued*)

(*continued*)

Author(s), Year	Title of Study	Objective	Methodology	Key Findings
L. Jaikala Karthikeyan (2023)	A Study on Customer Awareness on Cyber Security and Digital Payment Fraud Prevention in Nationalized Banks	To assess customer awareness regarding cybersecurity and fraud prevention in nationalized banks	The research is qualitative and exploratory, focusing on assessing customer knowledge and providing recommendations for enhancing digital payment security	Limited awareness among customers regarding cyber threats and fraud prevention measures
Surya et al. (2023)	Protecting Online Transactions: A Cybersecurity Solution Model	To propose a cybersecurity solution model for online transactions	This research employs a cybersecurity solution model approach to protect online transactions	Emphasizes OTP verification and password protection; highlights various cybersecurity threats in e-commerce
Sekhar & Kumar (2023)	An Overview of Cyber Security in Digital Banking Sector	To review cyber-attacks in the banking sector and propose cybersecurity measures	Qualitative analysis of cyber threats and security challenges in digital banking	Identified common cyber threats (phishing, spoofing, malware) and proposed integrated security solutions
Anitha Rathna et al. (2023)	Cyber Crime and Digital Payments in India: A Comprehensive Analysis	To analyze the causes, dangers, and potential solutions for cyber-attacks in digital payment systems	Comprehensive literature review and analysis of cybercrime trends in digital payments	Increasing risk of cyberattacks due to rising digital payments, need for cybersecurity measures, various types of cybercrimes explained

(*continued*)

(*continued*)

Author(s), Year	Title of Study	Objective	Methodology	Key Findings
Olifirov et al. (2023)	Cybersecurity Measures of the Digital Payment Ecosystem	To explore cybersecurity measures for digital payment systems, focusing on the digital ruble	Critical review of literature, analysis of cybersecurity measures, and reference to Global Cybersecurity Index (GCI)	Identified key cybersecurity challenges in digital payment systems; proposed technical, legal, and organizational measures for ensuring cybersecurity of the digital ruble platform
Suhasini Verma & Jeevesh Sharma (2023)	Mounting Cases of CyberAttacks and Digital Payment	To study the increasing cases of cyber-attacks in digital payment systems, explore their causes, and propose solutions through cybersecurity techniques	Literature review, statistical analysis of cybercrime incidents in India, and discussions on cybersecurity strategies	Rapid increase in digital payment fraud; key threats include phishing, malware, and identity theft. Solutions proposed include firewalls, MFA, and digital signatures
Vishnupriya et al. (2024)	A STUDY ON CYBER SECURITY ISSUES AFFECTING ONLINE BANKING AND TRANSACTIONS	To examine cybersecurity risks affecting online banking and transactions and assess current security practices	Primary data collected via questionnaires and secondary data from reports, research, and case studies	Identified phishing, malware, and identity theft as the most common threats; suggested multi-factor authentication (MFA) and user education as solutions

(*continued*)

(*continued*)

Author(s), Year	Title of Study	Objective	Methodology	Key Findings
Vishal Ramaiya et al. (2024)	Obstacles in the Way of Digital Payment: An Analytical Study	To identify the challenges and opportunities in digital payment systems in Rajkot district	Convenience-based sampling of 104 participants; data collected via surveys and interviews	Significant reliance on cash transactions due to lack of awareness and trust in digital payment methods
Juneja, Goswami, & Mondal (2024)	Cyber Security and Digital Economy: Opportunities, Growth and Challenges	To explore the dynamics of the digital economy's growth and the cybersecurity risks it faces	Literature review of past studies and analysis of global cybersecurity frameworks	Digital economy presents opportunities but also faces growing cybersecurity risks from AI, blockchain, and IoT; robust cybersecurity measures are essential for growth

Comparison of Digital Payment Systems in Terms of Security, User Adoption, and Key Features

Parameter	UPI	E-Wallets	Credit/Debit Cards	Net Banking
User Adoption	High, widely used for instant payments, especially among younger users	Moderate, primarily favored by tech-savvy younger users	Broadly used for both in-person and online transactions	Moderate, typically used for large payments and bill transactions
Security Concerns	High risk from phishing and MitM (Man-in-the-Middle) attacks	Moderate risk, mainly vulnerable to malware and data theft	Low to moderate, risk from card fraud like cloning	Moderate risk, primarily vulnerable to phishing and account breaches
Trust Levels	High, trusted due to regulatory backing and real-time confirmations	Moderate, some concerns over data privacy and security	High, trusted for long-standing usage but affected by occasional fraud	High, trusted for secure banking transactions via established institutions

(*continued*)

(*continued*)

Parameter	UPI	E-Wallets	Credit/Debit Cards	Net Banking
Transaction Volume	Very high, frequently used for small to mid-size payments	Moderate, used mainly for online shopping and services	High, commonly used for everyday purchases and larger transactions	High, mostly for larger transactions, utility bills, and secure payments
Fraud Incidents	Moderate, phishing and unauthorized access are common risks	High, frequent malware and data breach incidents	Low to moderate, issues with card theft or unauthorized transactions	Moderate, risks of phishing and unauthorized account access
Ease of Use	Very easy, integrated into mobile apps for quick transactions	Easy, but requires recharges and linking with bank accounts	Simple to use, though physical cards are necessary for offline use	Moderate, requires login credentials and additional security steps like OTPs
Cost of Transaction	Low, typically no extra charges for users	Low to moderate, some fees may apply for wallet transfers	Moderate, transaction fees may apply for specific types of card use	Low to moderate, occasional charges for inter-bank or cross-border transactions
Demographic Preference	Popular across age groups, especially among tech-savvy users	Preferred by younger, digital-native individuals	Broadly used across all age groups, particularly older users for offline payments	Favored by middle-aged to older users, typically for larger and more secure transactions
Key Security Measures	Multi-factor authentication (MFA), real-time monitoring	Encryption, biometric authentication, firewall protection	PIN, chip-and-PIN, OTPs for online use, encryption	Multi-factor authentication (MFA), OTPs, secure logins, encryption protocols

7 Research Methodology

Research Design
The study was conducted using an online survey through a Google Forms questionnaire, adopting a descriptive and exploratory approach. The questionnaire included a combination of open-ended and closed-ended questions, along with checkboxes to capture various types of responses.

Population and Sample Size
The research focused on respondents from the Rajkot district, with data collected from a total of 110 participants. And the data presented in this study, including the statistics and demographic breakdowns shown in Tables 1, 2, 3, 4, 5 and 6, were collected by the author through this survey.

Sampling Method
The study employed a convenience cum judgment sampling technique to select a suitable sample of users for data collection.

Tools for Analysis

- **Basic Analytical Tools:** To validate the data, the researcher used simple tools such as tables, figures, and percentages.
- **Statistical Analysis Tools:** Standard statistical metrics were applied, including the mean, standard deviation, variance, correlation, p-value, and chi-square test.

The data of domestic payment fraud underscore a troubling pattern about the frequency and consequences of fraudulent behaviour in the digital payment's environment. The increasing magnitude and monetary worth of fraudulent transactions indicate a rising cybersecurity threat associated with digital payments.

The research demonstrates that with the rise of digital payment methods, the complexity of fraudulent activities also increases. Therefore, it is crucial for carriers, including financial institutions and consumers, to prioritize cybersecurity measures. This underscores the significance of implementing robust security processes. To adequately mitigate these risks, it is crucial to deploy sophisticated fraud detection technologies, enhance user education, and promote industry collaboration. In general, the results emphasize the need of monitoring and adaptability in response to changing cyber risks in digital payment systems.

Demographic Factors Represented in Percentages
(See Figs. 1, 2, 3, and 4).

Figure 1 shows the demographic breakdown of survey participants based on their gender. Based on a pie chart, the gender distribution of 110 respondents is as follows: 52.7% are female and around 47.3% are male.

Figure 2 shows the age distribution of survey respondents is displayed. The graph illustrates that majority of the population falls within the age range of 21 to 30, with a significant proportion of persons below 21 years old. The proportions of older age groups

Table 1. "Statistics on Domestic Payment Frauds: Volume, Value, and Fraud Rates"

Domestic Payment Frauds				
	Volume (Lakh)	Value (Crore)	One in every X payment transaction fraudulent	FTS (Fraud / Payment Value * 10000)
September 2022	1.71	249	58177.41	0.127 bps
October 2022	1.79	220	59533.35	0.127 bps
November 2022	2.06	257	50620.69	0.143 bps
December 2022	1.54	204	72426.75	0.103 bps
January 2023	1.57	195	71681.44	0.106 bps
February 2023	2.29	317	47018.00	0.177 bps
March 2023	2.25	333	53907.48	0.142 bps
April 2023	1.75	273	68927.36	0.152 bps
May 2023	2.03	285	63051.71	0.147 bps
June 2023	1.74	265	72554.96	0.128 bps
July 2023	2.24	286	59898.87	0.146 bps
August 2023	2.40	320	58391.71	0.157 bps
September 2023	2.52	366	55057.76	0.174 bps
October 2023	2.23	335	66950.05	0.164 bps
November 2023	2.57	428	57618.81	0.209 bps
December 2023	2.92	432	53327.09	0.187 bps
January 2024	2.69	435	56335.82	0.207 bps
February 2024	2.53	503	59915.32	0.239 bps
March 2024	2.57	471	67394.38	0.177 bps
April 2024	2.38	414	70681.81	0.191 bps
May 2024	2.81	545	62637.44	0.240 bps

(Source: RBI. July 2024).

are reduced, indicating that the questioned population has a demographic composition that is mostly younger.

Figure 3 shows the educational qualifications of the participants in the survey: These statistics offer a thorough insight into the educational history of the respondents, revealing a diverse range of educational backgrounds within the studied population. Graduates constitute the largest proportion of responders, with postgraduates following closely behind. Fewer respondents have attained schooling up to the 12th standard, studied alternative professional courses, or were educated to the 10th standard or lower.

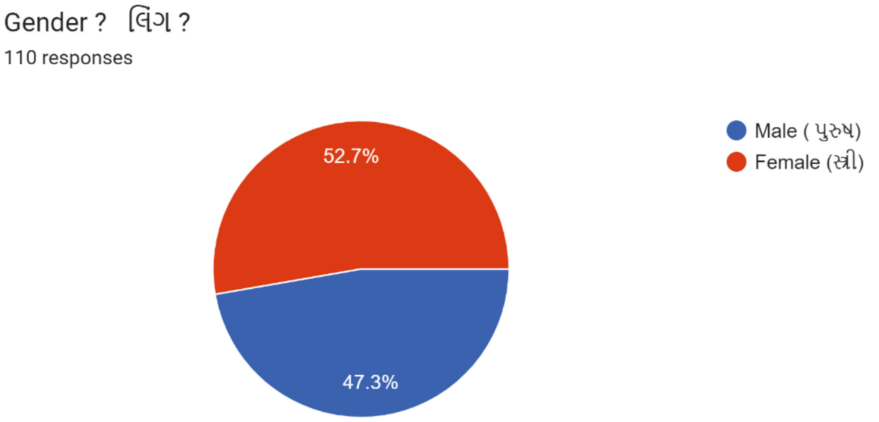


Fig. 1. Gender distribution of survey respondents.

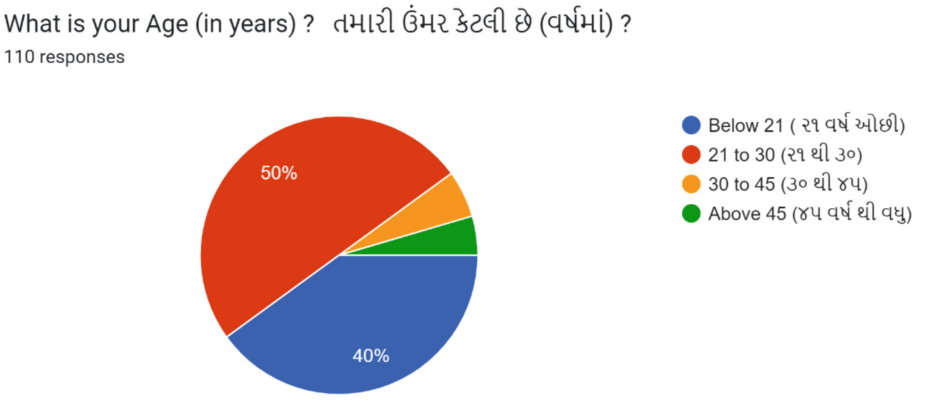


Fig. 2. Age-wise distribution of survey respondents.

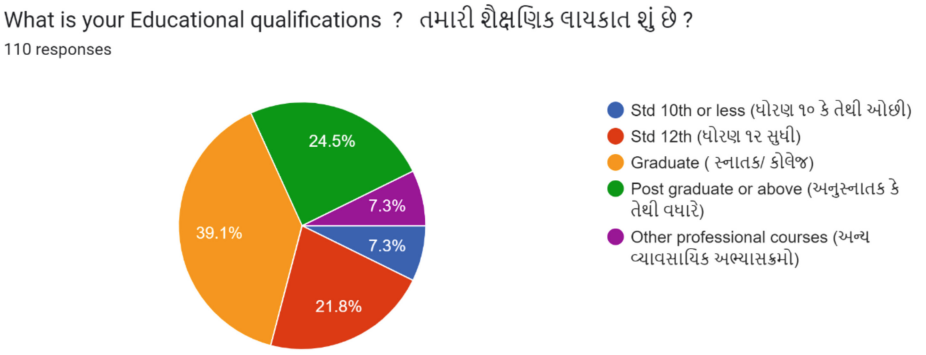


Fig. 3. Educational qualifications of survey respondents.

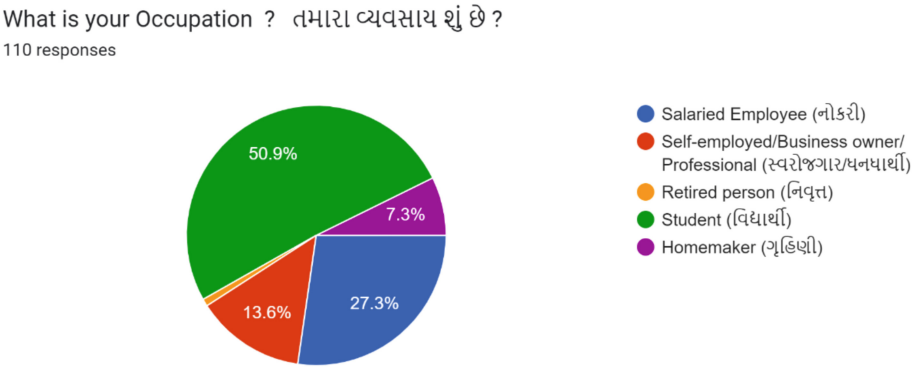


Fig. 4. Occupation-wise distribution of survey respondents.

Figure 4 shows the occupational distribution of those who utilize digital payment services. This data provides significant insights into the vocational diversity of consumers who use digital payment services. This highlights the notable prevalence of students and paid workers as actively engaged users of digital payment systems, together with a segment of self-employed persons and homemakers. The lack of participation of retired persons in this poll indicates a possible disparity in the acceptance of digital payments within this specific group.

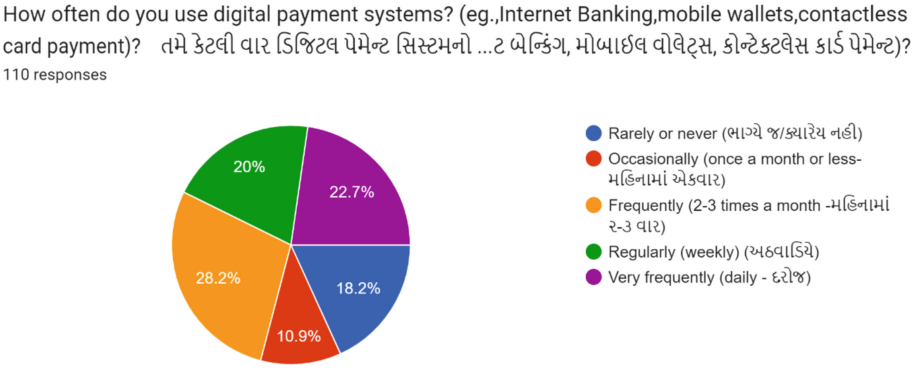


Fig. 5. Frequency of digital payment usage among respondents.

Figure 5 shows the data reveals a diverse range of digital payment usage frequencies among the respondents. While a majority of the respondents (71%) use digital payments frequently, regularly, or very frequently, there is still a significant portion (29.1%) who use them occasionally or rarely. This indicates varying levels of adoption and trust in digital payment systems, which may be influenced by factors such as demographic characteristics, familiarity with technology, and concerns about security.

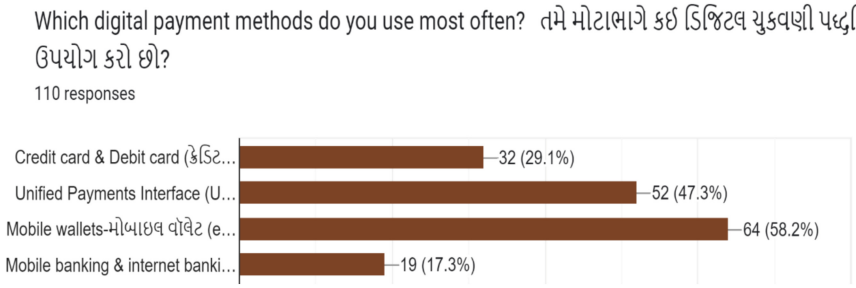


Fig. 6. Preferred digital payment methods among respondents.

Figure 6 shows that mobile wallets are the most preferred digital payment method among respondents, with 58.2% using them regularly. UPI follows with 47.3% of respondents favoring it for their transactions. Credit and debit cards remain popular, used by 29.1% of respondents, while mobile and internet banking are the least preferred, with only 17.3% opting for these methods. This figure highlights the growing preference for more convenient and quick digital payment options.

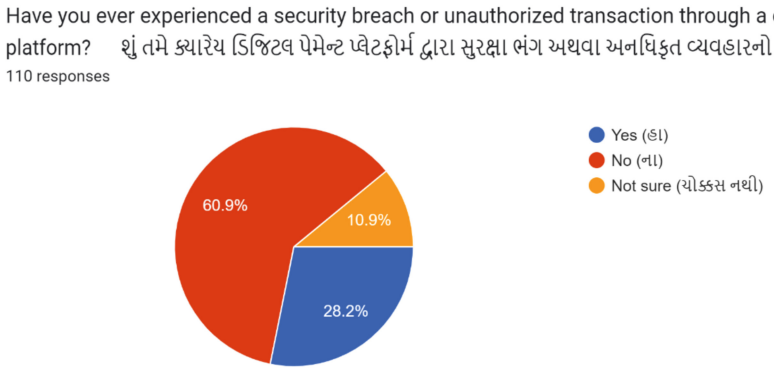


Fig. 7. Percentage of respondents who have experienced a security breach.

Figure 7 shows that 28.2% of respondents have experienced a security breach or unauthorized transaction through a digital payment platform, indicating a notable concern among users. Meanwhile, 60.9% of respondents reported no such experiences, suggesting that the majority find digital payments relatively secure. Additionally, 10.9% of respondents are unsure if they have encountered a breach, pointing to potential gaps in awareness or understanding of security issues.

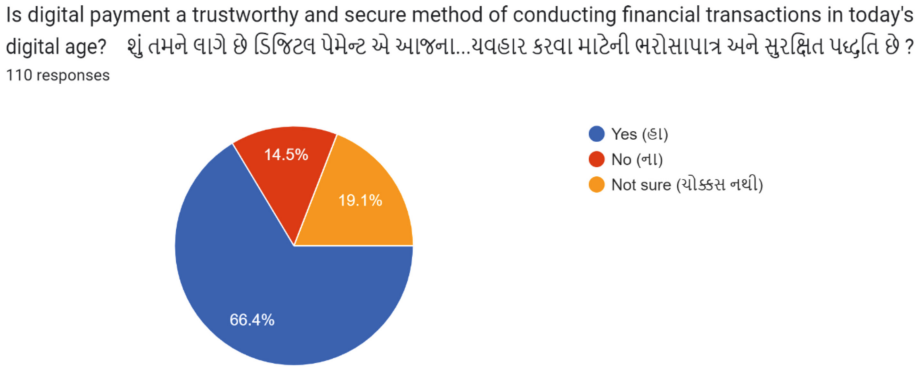


Fig. 8. Trust levels in digital payment security among respondents.

Figure 8 shows that 66.4% of respondents believe digital payment is a trustworthy and secure method for conducting financial transactions in today's digital age. This majority indicates strong confidence in digital payment systems. However, 14.5% of respondents do not share this trust, reflecting some scepticism or concerns about security. Additionally, 19.1% of respondents are unsure, suggesting that a significant portion of users may need more information or assurance regarding the safety of digital payments.

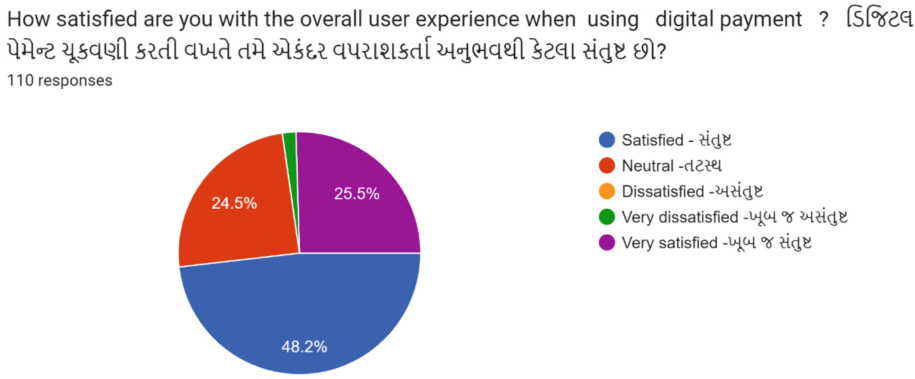


Fig. 9. Overall user satisfaction with digital payment systems.

Figure 9 shows that the majority of respondents have a positive user experience with digital payments, with 25.5% being very satisfied and 48.2% satisfied. A significant portion, 42.5%, remains neutral, indicating neither strong satisfaction nor dissatisfaction. Interestingly, dissatisfaction levels are extremely low, with no respondents feeling dissatisfied and only1.8% reporting being very dissatisfied. This data suggests that most users find digital payment systems generally reliable and user-friendly, though there is room for further improvement to convert neutral users into satisfied ones.

Hypothesis Testing
(See Table 2).

Table 2. Trust in Digital Payment Systems Among Users Who Have Experienced Security Breaches

Trust/security breaches	Trustworthy	Not Trustworthy	Not Sure	Total
Experienced Breach	17	6	11	34
No Breach	48	9	10	67
Not Sure	7	1	1	9
Total	72	16	22	110

Source: Data collected by the author (2024).

1. **Null Hypothesis (Ho):** There is no significant relationship between experiencing a security breach and the trust in digital payment systems.
   Alternative Hypothesis (H1): There is a significant relationship between experiencing a security breach and the trust in digital payment systems.

Statistical Test Results
The Chi-Square test yielded a value of 13.50 with 4 degrees of freedom and a P-value of 0.009. Since the P-value is less than the significance level of 0.05, we reject the null hypothesis. This indicates a statistically significant relationship between experiencing a security breach and the perceived trustworthiness of digital payment systems. Therefore, the results suggest that users who have experienced security breaches are indeed less likely to trust digital payment systems.

Table 3. User Satisfaction with Digital Payment Systems in Relation to Perceived Security and Trustworthiness

Satisfaction/trust	Very Satisfied	Satisfied	Neutral	Dissatisfied	Very Dissatisfied	Total
Trustworthy	24	36	10	0	2	72
Not Trustworthy	3	5	9	0	0	17
Not Sure	1	12	8	0	0	21
Total	28	53	27	0	2	110

Source: Data collected by the author (2024).

2. **Null Hypothesis (Ho):** There is no significant relationship between user satisfaction with digital payment systems and the perceived security and trustworthiness of these systems.

Alternative Hypothesis (H1): There is a significant positive relationship between user satisfaction with digital payment systems and the perceived security and trustworthiness of these systems.

Statistical Test Results

Chi-Square test statistics of 18.5 degrees of freedom, with a P-value of less than 0.05. This is an indication of a positive significant statistical correlation between user satisfaction of the digital payment systems and perceived security and trustworthiness of the systems. Thus, we fail to accept the null hypothesis and consider the alternative hypothesis. That is to say, customers who perceive digital payment as safe and reliable hold more satisfaction with the overall user experience.

Table 4. Gender-Wise Distribution of Security Breaches in Digital Payment Systems

Security breaches/ Gender	Experienced Breach	No Breach	Not Sure	Total
Female	13	38	7	58
Male	17	28	7	52
Total	30	66	14	110

Source: Data collected by the author (2024).

Table 5. Age-Wise Distribution of Security Breaches in Digital Payment Systems

Security breaches/ Age	Experienced Breach	No Breach	Not Sure	Total
Below 21	13	24	7	44
21 to 30	15	34	6	55
30 to 45	1	5	0	6
Above 45	1	3	1	5
Total	30	66	14	110

Source: Data collected by the author (2024).

3. **Null Hypothesis (Ho):** There is no significant relationship between demographic factors (such as gender, age, and educational qualification) and experiencing a security breach or unauthorized transaction through a digital payment platform.

   Alternative Hypothesis (H1): There is a significant relationship between demographic factors (such as gender, age, and educational qualification) and experiencing a security breach or unauthorized transaction through a digital payment platform.

Statistical Test Results

Chi-Square tests were done regarding different demographics for testing the null hypothesis. The null hypothesis cannot be rejected for gender, indicating no variation in concern

Table 6. Education-Level Distribution of Security Breaches in Digital Payment Systems

Security breaches/education	Experienced Breach	No Breach	Not Sure	Total
Std 10th or less	0	6	2	8
Std 12th	7	14	3	24
Graduate	13	24	6	43
Post graduate or above	9	17	1	27
Other professional courses	1	6	1	8
Total	30	67	13	110

Source: Data collected by the author (2024).

for cyber security, while regarding age and educational qualifications, variation has been detected and the null hypothesis has been rejected on the basis of it. From the above, it is clear that the demographics of age and education influence perceptions of cyber security in the digital payment system.

Future Scope

1. **Advanced Fraud Detection and Prevention:** As the complexity of cyberattacks increases, future research areas could encompass the integration of AI and ML models in predicting and preventing fraud more efficiently. This comes in the form of real-time analytics on transactional data for anomaly detection and at the time of detecting anomalies, must respond to potential threats with a lot of speed.
2. **User-centric security solutions:** In light of these variations in the level of trust of different demographic groups, further research may be conducted on personal security mechanisms that make provision according to the needs and behavior of specific user groups, such as geriatric individuals or less technologically savvy users. Similarly, tailored cybersecurity awareness solutions may be developed for different user groups as well.
3. **Cross-regional and cross-cultural comparison:** An extension of this study to a wider geographical area would enable its understanding of how cultural and regional factors dictate or influence cybersecurity risks and user trust. A comparative study across countries or regions will be helpful in pinpointing the general challenges identified universally versus those unique only to a particular region in digital payment systems.
4. **Blockchain and Cryptography-Based Security:** Possible Future Work The next and significant area of research could be on how blockchain technology combined with advanced cryptography techniques can be integrated to enhance the security level for digital payment systems. The platforms are likely to eliminate usual threats of identity thefts and data breaches due to their decentralized nature.
5. **Regulatory Frameworks and Policy Formulation:** As the digital payment systems are still in the growth phase, governments or regulatory bodies have turned out to

be very crucial players. Emerging Regulations: Some of the upcoming regulations may include data protection law and cyber security standards. Future work would be focused on how it impacts the security and trust of the digital payment platforms.

6. **Longitudinal Studies of Trust Dynamics:** This work finally brings some light to current user trust, and future work could conduct longitudinal studies on other significant patterns - how trust over time changes concerning newly implemented cybersecurity measures or evolving cyberattacks. A longitudinal study can evolve into a deeper understanding of the dynamics of building or eroding trust over time.
7. **Cyber Security in Emerging Payment Technologies:** As newer digital payment forms, including biometric authentication, wearable devices, and IoT-based transactions, begin to gain traction, there arises a need for more research in these domains, as well as their specific issues related to cybersecurity.

Findings of the Study

1. **Consequences of Security Breach of Trust:** Almost 28.2% of participants responded affirmatively to the claim that they have undergone a security breach or unauthorized transaction with the aid of digital payment mechanisms. The statistical test results show that there is a direct relationship between security breaches and reduced trust in the use of digital payment mechanisms. This further implies that users who have experienced security breaches are relatively less likely to trust these mechanisms.
2. **Increased cybersecurity threats:** The research reveals an upward trend in cybersecurity threats as digital payment systems are used more frequently. These include identity theft, phishing, malware, and ransomware attacks. Such attacks are becoming more frequent and sophisticated, posing a significant threat to the integrity and security of digital transactions.
3. **User Satisfaction and Perceived Security:** Overall, this study shows that user satisfaction is indeed related to the perceived security and trustworthiness of digital payment platforms. The perception regarding security and trustworthiness is significantly higher for users who feel that digital payment platforms are secure and trustworthy, compared to those who have concerns about security.
4. **Demographic Factors Impacting Perception of Cybersecurity of Digital Payments:** Age and educational background are two major demographic factors that influence users' perceptions about cybersecurity risks in digital payment systems. The better educated a user is, the more perilous the perceived potential risks would be, and consequently, there is a greater likelihood that such users would be inclined to undertake precautionary measures than would older users or less-educated users.
5. **Preferred digital payment methods**: A study finds that, on the whole, mobile wallets and the Unified Payment Interface (UPI) formed the most preferred payment methods for users. A total of 58.2 percent of people use mobile wallets, while 47.3 percent of the respondents Favor UPI. Users prefer these apps due to their speed and high convenience, although some remain concerned about security.
6. **Need for Enhanced Security Measures**: The findings underscore the need for robust security measures to address the growing risks associated with digital payments. Financial institutions should invest in advanced fraud detection systems, implement

multi-factor authentication, and enhance user education on cybersecurity practices. These measures are critical to ensuring the safety and trustworthiness of digital payment platforms.

Approaches to Enhancing the Security in Digital Payment Systems

1. **Enhancement in Cyber Security Awareness:** The consumers have to be abreast with the emerging cyber security threats and take due and proper adequate safeguards so that the consumer information regarding finance may remain safeguarded.
2. **Regular Security Audit:** The financial institutions are at regular periods undertaking audits and vulnerability testing of their digital infrastructures. This helps to pre-identify the potential gaps in security before they are actually exploited.
3. **Encourage the deepened adoption of multi-factor authentication:** in all electronic transactions, Through the use of multiple levels of verification, this model substantially secures access and makes any illegal access much more challenging.
4. **Investment in State-of-the-Art Fraud Detection Technology:** The regulator should encourage financial institutions to invest in fraud detection systems that are at least sophisticated enough to apply Artificial Intelligence and Machine Learning principles for the detection of suspicious activities, assuring an effective response in real time.
5. **Stakeholder cooperation:** Enable greater cooperation between the financial sector, the government, and cybersecurity experts regarding developing threats and ways to effectively neutralize them. This can, in turn, be considered a panacea for enhancing overall security in the arena of digital payment.
6. **Establishment of Mechanism for User Feedback**: There should be strong feedback to garner useful inputs from the users regarding experiences and concerns related to digital payments' security. Feedback received can usefully guide improvements and make sure that security measures are according to the expectations of the users.
7. **Policymakers need to come up with and implement laws**: that stipulate the minimum essential requirements in cybersecurity for digital payment systems. This would ensure the adherence of service providers to the best practices toward ensuring information and transactions of customers are secured.
8. **Encourage Continuous Research and Development:** Allow the continuous research and development of cybersecurity technologies, particularly those touching on the digital payment systems. For integrity and security in these systems, it is very important that continuous updates always try to outpace the growth of threats.
9. **Explain in detail the comprehensive incident response plans**: that shall be spelled out to describe exactly what will be done in the event of a security breach. The plans must include proper strategies for timely communication with the consumers affected to prevent further harm.
10. **Design Security Features with the Needs of Users in Mind:** Impress during the design phase that a digital payment interface secures users by making security features available to them and understandable to them. Such steps shall instil confidence in users that they are safe while using the Digital Payments system.

8 Conclusion

This paper contends for the vital role of enhancing cybersecurity inside the rapid advancing discipline of virtual payments. The developing incorporation of virtual fee systems into normal transactions leads to a corresponding growth in safety vulnerabilities, emphasising the need for an intensive and specific protection policy. The analysis yields many giant findings:

1. **Enhancing User Awareness:** It is vital to enhance clients' knowledge of ability cybersecurity threats and the significance of enforcing sturdy preventive measures to sufficiently shield their financial statistics.
2. **Implementing Strategic Measures:** The paper shows several strategies to improve the safety and reliability of virtual fee protocols. These steps consist of the normal upgrading of safety protocols, good sized costs in modern cybersecurity technology, and the supply of continuous education to customers on moral virtual commercial enterprise practices.
3. **Establishing and Maintaining Trust:** Improving the safety framework of virtual bills is important for promoting and shielding consumer trust, which is a basic foundation for the ongoing expansion and recognition of virtual economic services.

In order to successfully cope with the evolving intricacy of cybersecurity in virtual payments, it is vital for all stakeholders inside the gadget to cooperate and combine their efforts. By collaboratively enforcing the advocated movements, it's far viable to create a stable and reliable environment that fosters consumer self-belief and enables the in-addition boom of virtual payment networks.

References

Patil, S.: A cybersecurity agenda for India's digital payment systems, pp. 8–33. Gateway House: Indian Council on Global Relations (2019)

Gs, A.P., Arivazhagan, R.: A study on user perceived trust in mobile wallet. Int. J. Manag., 125–133 (2021)

Bhupali Shah, D.S.: A review of cyber security issues in online banking and online transactions. NeuroQuantology, 405–418 (2022)

Moon, I.T., Shamsuzzaman, M., Mridha, M.M.R., Rahaman, A.S.M.M.: Towards the advancement of cashless transaction: a security analysis of electronic payment systems. J. Comput. Commun., 103–129 (2022)

Jeevesh Sharma, S.V.: Mounting cases of cyber-attacks and digital payment. In: Cybersecurity Issues, Challenges, and Solutions in the Business World, pp. 60–80 (2023)

Karthikeyan, L.J.: A study on customer awareness on cyber security and digital payment fraud prevention in nationalized banks. J. Harbin Eng. Univ., 459–472 (2023)

Kumar, M.: An overview of cyber security in digital banking sector. East Asian J. Multidisc. Res., 43–52 (2023)

Anitha Rathna, G.D.M.: Cyber crime and digital payments in India: a comprehensive analysis. In: Vision for Vishwa Guru India: Initiatives for Global Leadership by 2047, Pollachi, Tamil Nadu, pp. 1–7 (2023)

Surya, S.R.: Protecting online transactions: a cybersecurity solution model. In: International Conference on Advance Computing and Innovative Technologies in Engineering, pp. 2630–2634 (2023)

Vishal Ramaiya, N.K.: “Obstacles in the way of digital payment” – an analytical study. In: ASCIS 2023, Rajkot, pp. 309–325 (2024)

Reserve Bank of India - Annual Report (rbi.org.in). Accessed 28 July 2024

Olifirov, A., Makoveichuk, K. A., Petrenko, S.: Cybersecurity Measures of the Digital Payment Ecosystem. In: CEUR Workshop Proceedings, vol. 2914, pp. 234−247 (2023). https://ceur-ws.org/Vol-2914/paper20.pdf

Vishnupriya, N., Tanusha, P., Bhujanga Rao, P.: A study on cyber security issues affecting online banking and transactions. Int. J. Adv. Res. Innovative Ideas Educ. **9**(6), 1655−1665 (2024). Available at:https://www.researchgate.net/publication/378869599.

Juneja, A., Goswami, S. S., Mondal, S.: Cyber security and digital economy: opportunities, growth and challenges. J. Technol. Innovations Energy **3**(2), 1−22 (2024). https://doi.org/10.56556/jtie.v3i2.907

Enhanced Web Browser Forensics: Innovative Methodologies for Evidence Collection and Analysis

Nishant Navinbhai Joshi(✉) and Sunil Lalchand Bajeja

Faculty of Computer Application, Marwadi University, Rajkot, Gujarat, India
nishantjoshi33@gmail.com

Abstract. Web browsers are widely used computer applications that assist user in engaging in various activities, such as browsing the internet, downloading files, interacting with social medial networks, and managing email. As web browsers store critical user data, they become essential for investigating online malicious activities. This study explores various method used to analyze web browser data in the field of digital forensics, focusing on different types of data, including url's, timestamp, browser versions, download files and search queries. The organization of this data depends on the operating system, which requires specific methods tailored to each situation. Digital forensics, a crucial branch of forensic science, focuses on extracting and examining digital information from devices to detect criminal activities or verify device breaches. Web browser analysis plays a significant role in digital forensics, contributing to both criminal and civil cases due to the vast amount of evidence generated by online activities. This study reviews the primary techniques and methods used for web browser analysis, assessing their advantages and disadvantages. Additionally, it presents an innovative approach to evidence collection and analysis to advance forensic science. As web browsers enable tasks like internet browsing, file downloading, email access, and social media interaction, they are vital sources of evidence in digital forensic investigations, especially in cases involving cybercrime or malicious activities. This research focuses on the methods used to analyze key types of web browser data, such as URLs, timestamps, browser versions, downloaded files, access times, and search queries. Given the variation in data storage methods across different browsers and operating systems, forensic techniques must be customized for each scenario. By reviewing the current web browser analysis tools and techniques, this study highlights their advantages and limitations while introducing a novel framework for evidence collection and analysis. This framework aims to improve the accuracy and effectiveness of forensic investigations, addressing the complexities of browser data across platforms and advancing the field of forensics science in the investigation of digital crimes.

Keywords: Digital forensics · browser forensics · Evidence collection · types of browsers · Forensics Tools

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 139–164, 2025.
https://doi.org/10.1007/978-3-031-86305-9_11

1 Introduction

Internet security has been a significant and escalating problem for numerous years. Internet security can be jeopardized not just by malware, fraud, system breach, or destruction, but also through the monitoring of online behavior. To counter these dangers, data encryption is now a standard default configuration. Firewalls and antivirus software are crucial instruments in combating cybercrime. Internet web browsers are routinely utilized by the majority of persons and are present on computers, mobile devices, gaming consoles, smart televisions, wristwatches, automobiles, and household appliances. Web browsers retain user data, including visited websites and the corresponding date and time of internet searches [1]. Besides erasing their browsing history, users can also inhibit the retention of such data by utilizing 'private browsing' functionalities and tools. [2] Users may opt for private surfing features and capabilities for various purposes, such as online gift purchasing, website testing and debugging, and utilizing public computers. In the contemporary digital era, web browsers have emerged as essential instruments for accessing and engaging with the internet, rendering them crucial sources of digital evidence in forensic inquiries. With the rise of cybercrimes and online activities, web browser forensics has advanced significantly, providing more refined techniques for analyzing browser artifacts, monitoring user behavior, and recovering essential data for investigations. Recent breakthroughs, including refined browser fingerprinting methods, augmented cache and cookie analysis, and the capability to extract metadata from encrypted web sessions, have considerably bolstered the capacity to attribute online behaviors to particular people or devices. The use of machine learning algorithms into forensic tools has facilitated more precise identification of anomalies in browsing habits, yielding insights into harmful behaviors including phishing, fraud, and data exfiltration. Furthermore, contemporary browsers utilize an array of privacy-enhancing functionalities, like encrypted DNS queries and incognito modes, which pose additional hurdles for forensic investigators. Advancements in network traffic analysis, artifact recovery from browser extensions, and the creation of tools capable of circumventing certain privacy measures have mitigated these issues, enhancing the robustness of browser forensics in digital investigations. The related work section is thorough but could benefit from more critical analysis rather than just summarization.

1.1 Digital Forensics Life Cycle

The initial phase of the digital forensics' lifecycle is identification. Digital forensics is an essential discipline in the contemporary technology-centric landscape, cybercrime and digital evidence are increasingly common. The digital forensics life cycle is a systematic method for the investigation and analysis of digital evidence, maintaining integrity and reliability throughout the procedure. This blog will guide you through the fundamental steps of the digital forensics lifecycle, offering a through comprehension of each phase is as below (Fig. 1).

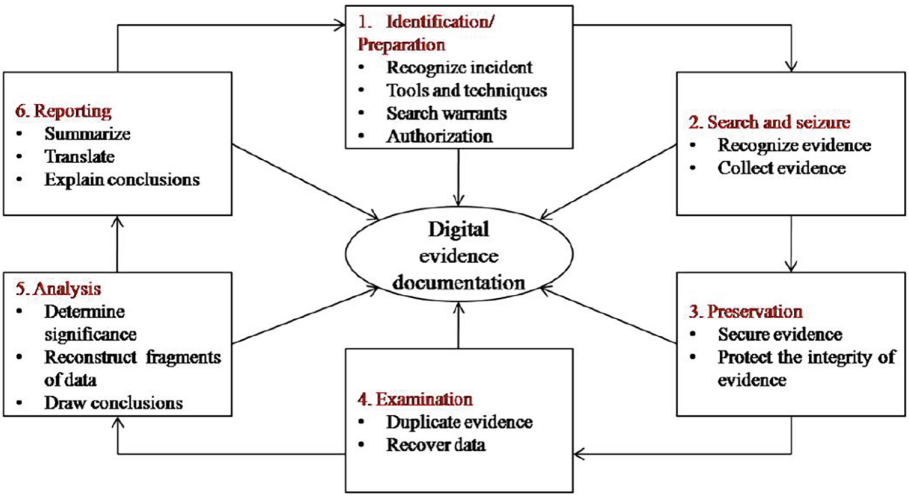


Fig. 1. Digital Forensics Life Cycle (https://www.startertutorials.com/blog/digital-forensics-life-cycle.html)

Digital Forensics Lifecycle Steps

A. Preparation
B. Collection
C. Storing & Transporting
D. Examination/Investigation
E. Analysis
F. Reporting
G. Testing

(1) **Preparation** The initial phase of the digital forensics lifecycle is identification. This phase entails identifying and assessing possible sources of digital evidence. In this phase, forensic investigators ascertain equipment, data, and pertinent information that may be useful to the inquiry. Prevalent sources encompass computers, mobile devices, servers, cloud storage, and network logs.

Essential Activities:

Examining the crime scene (physical or digital)
Identification of devices and data sources
Identifying potential evidence sites

Accurate identification guarantees that all pertinent evidence is documented, thereby averting the omission of any essential data. **Types of examples is as below.**

(1) Different files
(2) File and file system
(3) Process and files
(4) Log files

(2) **Collection**:- The collection phase entails the methodical acquisition of digital evidence from designated sources. This stage necessitates meticulous processing to guarantee that data is acquired in a forensically sound manner without modifying the underlying evidence.

Essential Activities:

(1) Retrieving data from many digital devices and storage media.
(2) Acquiring ephemeral data (e.g., RAM contents) prior to its loss.
(3) Meticulously documenting the collection process.
(4) Meticulous and exhaustive collecting guarantees the acquisition of all possible evidence, enabling a full study.

Ex. Digital evidence can be collected from many sources is as below

(1) Mobile phone
(2) Digital camera
(3) Hard drives
(4) CDs
(5) USB memory devices

Non-Obvious sources can be

(1) Digital thermometer setting.
(2) Black boxes inside automobiles.
(3) RFID tags

(3) **Storing:** - Upon the identification of potential evidence, the subsequent stage is preservation. This entails safeguarding and maintaining the integrity of digital evidence to avert manipulation, modification, or deletion. Chain of custody protocols are instituted to preserve a comprehensive record of individuals who managed the evidence and the corresponding timestamps.

Essential Activities:

(1) Generating incremental replicas (imaging) of digital evidence.
(2) Establishing chain of custody documentation.
(3) Employing write-blocking instruments to avert alterations to the original evidence.
(4) Preservation is essential for upholding the validity and admissibility of evidence in judicial processes.

(4) **Examination: -** During the examination phase, forensic professionals analyses the gathered evidence to discern pertinent facts. This entails employing specialized instruments and methodologies to retrieve deleted files, examine metadata, and reveal concealed data.

Essential Activities:

(1) Examining file systems and data architectures.
(2) Restoring lost, encrypted, or concealed data.
(3) Retrieving metadata for contextual analysis.

The examination step discloses essential information that may not be readily visible, revealing critical elements imperative to the investigation. Some tool which can create entire hard drive images are

(5) **Analysis**: - The analysis phase involves interpreting the extracted and analyzed data to derive significant findings. This entails comparing results with the investigative context, recognizing patterns, and reconstructing occurrences.

Essential Activities:

(1) Integrating data from several sources.
(2) Performing temporal analysis to reconstruct occurrences.
(3) Recognizing patterns and irregularities.
(4) Window registry inspection.
(5) Discovering and cracking passwords.
(6) Extracting email and image.

Analysis converts raw data into actionable intelligence, yielding insights that can advance the inquiry and bolster legal claims. The digital evidence must be analyzed to determine the type of information stored on it.

Example of forensics tools

(1) Forensics tool kit (FTK)
(2) Encase
(3) Scalpel (file carving tool)
(4) The sleuth Kit (TSK)
(5) Autopsy

Example Type of Digital analysis

(1) Media analysis
(2) File System analysis
(3) Application analysis
(4) Network analysis
(5) Image analysis
(6) Video analysis

(6) **Reporting**: - Upon completion of the analysis, a report is produced. The report may be presented orally, in writing, or in both formats. The report encompasses all details regarding the evidence in the analysis, interpretation, and attribution phases. Consequently, the findings in this phase should enable the confirmation or refutation of the allegations.

The report includes several general elements.

(1) Identity of the reporting agency
(2) Case identifier or submission number
(3) Case investigator
(4) Identity of the submitter
(5) Date of receipt
(6) Date of the report
(7) Enumerative catalog of items presented for evaluation
(8) Examiner's identity and signature
(9) Concise account of procedures undertaken throughout the assessment
(10) Findings/conclusions

(7) **Testing**:- This stage entails the presentation and cross-examination of expert witnesses. An expert witness may provide testimony in the following formats

(1) Testimony is founded on adequate facts or data.
(2) Testimony results from dependable principles and methodologies.
(3) The witness has consistently applied principles and methods to the case facts.

The court occasionally reprimands experts lacking sufficient knowledge. Precautions to observe when gathering digital evidence include: No actions undertaken by law enforcement agencies or their representatives should alter the evidence. To access the original data stored on a computer, an individual must possess the requisite competence.

An audit trail or other documentation of all processes applied to digital evidence must be established and maintained.

The individual overseeing the investigation bears ultimate responsibility for ensuring compliance with the law and relevant regulations.

2 Browser Forensics

A wide range of individuals, including those believed to be under investigation as possible suspects, use the Internet. Someone under suspicion may employ a web browser as a means to acquire information, obfuscate their illicit actions, or investigate novel approaches to perpetrating criminal acts. An integral component of digital forensic investigations typically involves the analysis of internet surfing activity traces. A wide range of activities performed by A suspect browsing the Web can lead to incriminating evidence being found on the computer, encompassing the act of information retrieval. As a result, the examination of a suspect's computer by an investigator can provide significant and relevant empirical evidence. After retrieving several types of data taken from a computer belonging to a suspect, including Temporary storage memory, browsing records, saved website information, and records of downloaded items, it is possible to analyze this evidence to ascertain the websites that were visited, the timing and frequency of access, and the search engine keywords utilized by the suspect. The analysis of web browser log data encompasses a range of research topics and technologies, which exhibit notable similarities in their respective properties. The aforementioned studies and tools have been specifically developed to cater to a particular internet browser or a particular log document associated with that window. There exists a diverse range of web browsers that currently provide internet services, enabling users to leverage and compare many web browsers concurrently. Hence, conducting individual examinations for each web browser is an insufficient approach for detecting evidentiary indications of a user's unlawful online behavior. Moreover, it is insufficient to solely analyze a solitary file hosted on a singular web browser, as the evidence may be dispersed throughout numerous files. The user's text is "[16]" (Fig. 2).

2.1 Different Types of Modes or Web Browsers are as Below

(1) **Portable mode/browser: -** This mode is only storing a directory, when this mode is activated, the browser Not storing browsing history, cookies, site data, or form inputs. Typically used to improve privacy by avoiding the retention of local data that could be accessed in the future. This mode is also known as "Incognito mode or private browsing mode".

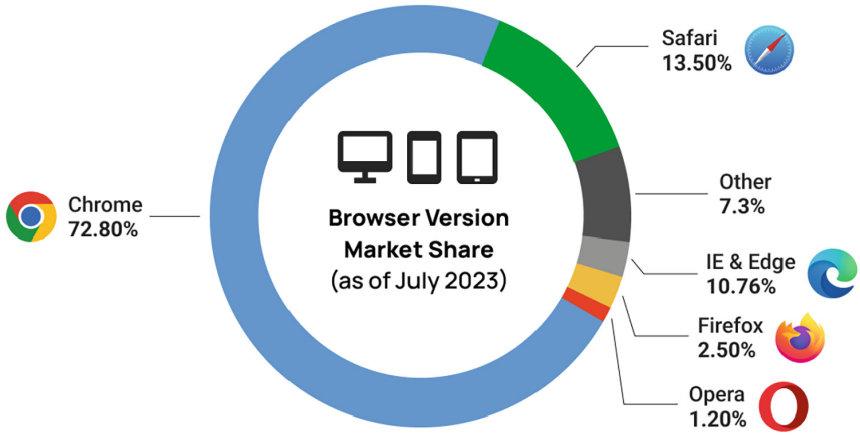


Fig. 2. Life cycle of Digital Forensics (URL: https://www.researchgate.net/figure/Digital-Forensic-Investigation-Lifecycle_fig1_304037254)

The name Of the Browser used in portable browsers is as below. (1) Tor Browser (2) Epic Privacy Browser (3) Mozilla Firefox (4) Google Chrome (5) Microsoft Edge (6) Apple Safari

(2) **Public Browser:** Private browser or private browsing modes in regular browsers are created to improve user privacy by avoiding storing browsing history, cookies, or other local data. Here are some well-known private browsers and the private modes available in popular web browser.

The name of the browser used in Public Browser is below. (1) Google Chrome (2) Opera Browser (3) Mozilla Firefox (4) Apple Safari (5) Microsoft Edge

(3) **Private Browser: -** A private browser or private browsing modes in standard browsers are specifically created to protect user privacy by restricting the storage of browsing history, cookies, and other local data. Here are some well-known private browsers and the private modes available in popular web browser.

The name of the Browser used in private browser is as below (1) Tor Browser (2) Brave Browser (3) Epic Privacy Browser (4) Apple Safari (5) Opera Browser.

Table 1. List of virus attacks on browser year 2020 to 2024 (URL = "https://www.indusface.com/blog/web-browser-based-attacks-how-to-protect-your-end-users/")

S.R.No	Browser Name	Year	Types of Attack	Infected things
1	Google Chrome	2020	Phishing Attack	Targeted Chrome users by creating a fake login page to steal credentials
2		2021	Malicious Extensions	Over 111 malicious extensions were discovered affecting millions of users by injecting and stealing data
3		2022	Zero-Day Exploits	Multiple zero-day vulnerabilities were exploited, requiring emergency patches to prevent widespread damage
4		2023	Drie by download	Increased instances of drive-by downloads from compromised websites, leading to malware infections
5		2024	Crypto jacking	Unauthorized cryptocurrency mining scripts were found running in the background of many users' browsers, leading to performance issues
1	Mozilla Firefox	2020	Cross-Site Scripting (XSS)	Exploited XSS vulnerabilities allowed attackers to inject malicious scripts into web pages viewed by other users
2		2021	Malicious Add-ons	Several malicious add-ons were discovered, leading to unauthorized data access and unwanted actions
3		2022	Phishing Attack	Increased phishing attempts targeting Firefox users through deceptive websites and emails
4		2023	Drive-by Downloads	Similar to Chrome, Firefox users faced drive-by downloads from compromised websites

(*continued*)

Table 1. (*continued*)

S.R.No	Browser Name	Year	Types of Attack	Infected things
5		2024	Zero-Day Exploits	Exploitation of previously unknown vulnerabilities required urgent patches
1	Microsoft Edge	2020	Browser Hijacking	Malicious software changed Edge's settings, redirecting users to unwanted sites and displaying unwanted ads
2		2021	Phishing Attacks	Numerous phishing campaigns targeted Edge users, especially through malicious emails
3		2022	Zero-Day Exploits	Exploited zero-day vulnerabilities led to significant security breaches
4		2023	Clickjacking	Users were tricked into clicking on deceptive elements, leading to unintended actions
5		2024	Drive-by Downloads	Increased instances of drive-by downloads affecting Edge users
1	APPLE SAFARI	2020	Malvertising	Malicious advertisements led to malware infections among Safari users
2		2021	Cross-Site Scripting (XSS)	Exploited XSS vulnerabilities led to unauthorized script execution
3		2022	Phishing Attack	Safari users were targeted by phishing campaigns designed to steal personal information
4		2023	Crypto jacking	Unauthorized cryptocurrency mining scripts ran in the background, affecting performance
5		2024	DNS Spoofing	Traffic was redirected from legitimate websites to malicious ones, leading to data theft

2.2 Analysis of Web Browser Records Types of Methods is as Below

(1) **Manual Analysis**

Direct Examination: Manually exploring browser folders to find relevant files.

Registry Analysis: Examining the registry for browser-related information (Windows).

(2) **Forensic Tools**

Commercial Suites: EnCase, FTK, and Cellebrite offer comprehensive browser forensic capabilities.

Open-Source Tools: Autopsy (part of The Sleuth Kit) provides free alternatives.

Specialized Browser Tools: Net Analysis, Browser History Examiner, Magnet IEF focus on browser data.

(3) **Database Analysis**

SQLite: Many browsers use SQLite databases. Tools like DB Browser for SQLite can examine them.

ESEDatabaseView: For browsers using Extensible Storage Engine (ESE).

(4) **File Analysis**

Hex Editors: Examining file contents in hexadecimal format.

String Extraction Tools: Extracting text strings from files.

2.3 Analysis of Web Browser Records

Web browsers are commonly used programs on electronic devices. Individuals participate in different online activities through web browsers on varying operating systems. Web browsers offer services like searching for information, sending emails, shopping online, reading news, banking online, using social media, and writing blogs [6]. This reasoning is considered one of the most essential aspects of evidence examination. Digital forensics experts use data like the websites accessed, search terms used, and the timing of these activities to determine the type of criminal activity. Furthermore, it is essential to evaluate the usage of multiple web browsers at the same time. There are many tools, both open source and licensed, that can be used for conducting analysis. Users often participate in the tasks described in Table 1 [2] through the use of web browsers (Table 2).

Table 2. Analysis of web browser records

Web Browser User Engagement	Keywords in URL	Description
Search	'Search', 'word' (related to search	Users search for information using keywords in search engines. The URL may contain terms related to the search query.

(*continued*)

Table 2. (*continued*)

Web Browser User Engagement	Keywords in URL	Description
Email-	'mail', 'e-mail'	Users access web-based email services to send and receive emails. URLs often include terms like "mail" or "e-mail"
Blog	'blog'	Users engage in blogging activities. Such as reading. Creating, or managing blog posts. URLs usually contain the term "blog".
Social media	'Facebook', 'Twitter', 'Instagram' etc.	Users interact on social media platforms. URLs typically include the name of social networks such as "Facebook", "Twitter", and "Instagram".
News	'news'	Users access online news portal to stay updated on portals to stay updated on current events. URLs generally contain the word "news".
Video, Visual content	'video'	Users watch videos or view visual content on various platforms. URLs typically contain the word "video".
Music	'music', 'MP3'	Users listen to music or download MP3 files. URLs often include terms like "music" or "mp3".
Banking	'bank' 'credit'	Users perform online banking activities, including managing accounts and trans

2.4 Related Work in Browser Forensics to Collected Evidence

"During examining online browser activity, a range of evidence has been identified, which are enumerated as follows".

1	**History**	Record the dates and times of websites visited, store data for each local user account, track how often sites are accessed, oversee entry to local system documents, study the timestamps of visited sites, and maintain pertinent data records.
2	**Cookies**	They provided information about the websites that were visited and the behaviors that may have occurred there. These are compact text files that store user data.

(*continued*)

(*continued*)

3	**Cache**	Provides the investigator with a momentary representation of the internet content that a user was seeing and determines the websites that have been accessed. Stored files are associated with a particular local user account; Timestamps indicate the initial saving and most recent viewing of the site. Displays the specific files that the user accessed on a particular website.
4	**Download**	The investigator would also inspect the default download folder, as it is where all the downloaded files are stored. Internet-acquired files.
5	**Book marks**	Archived URLs can offer valuable information on regularly visited or significant websites.
6	**Session**	Files refer to the data containing information on open tabs and sessions. These files have the potential to be retrieved even after the browser has been closed.

When a hacker assaults a browser, a researcher can utilize diverse sorts of proof specified in the table to examine and comprehend the hacker's activities. The browsing history documents the websites visited by the hacker, detailing the times and frequency of access, so facilitating the tracking of their online activities. Cookies provide essential insights into user behavior on certain websites, disclosing login sessions, preferences, and even compromised accounts. Through the examination of the cache, the investigator can acquire a snapshot of the content the hacker engaged with, encompassing any files or photos visited, as well as timestamps denoting the timing of these interactions.

The download folder may disclose files obtained by the hacker from the internet, including malware or compromised data. Examining bookmarks provides insights into the websites the hacker deemed significant or regularly accessed, potentially revealing their motives or connections to other unlawful activity. Ultimately, session files provide the retrieval of data on the hacker's active tabs and web sessions, even post-browser closure, providing investigators with a comprehensive understanding of the hacker's activities immediately prior to exiting or powering down the machine**.**

3 Related Tools in Browser Forensics

3.1 The Existing Tool that Identifies Evidence During a Cyber-Attack on the Web Browser is as Below

(1) **Auto spy: -** Autopsy is a software tool designed for conducting digital forensic investigations on both Windows and Linux operating systems. The tool offers capabilities for data carving, timeline analysis, and web artifact analysis. The autopsy retrieves the browsing history, stored cookies, and bookmarked websites from the Firefox, Chrome, and Internet Explorer web browsers. Law enforcement, military, and business examiners utilize it to conduct computer investigations. It is possible to utilize this tool to retrieve images from the memory card of your camera (Fig. 3).

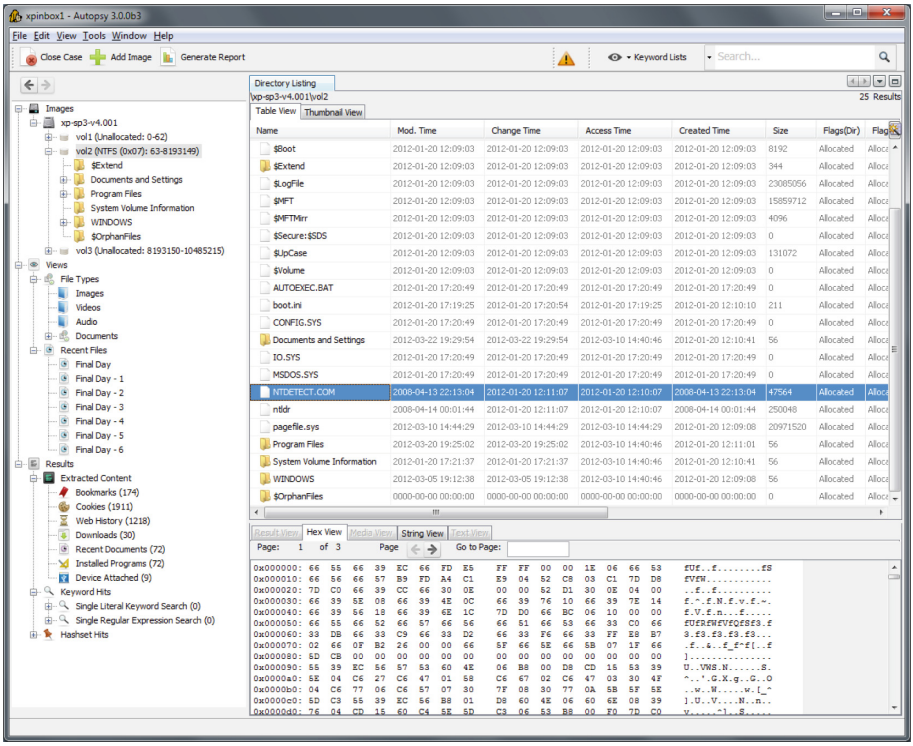


Fig. 3. Auto Spy tool using in browser forensics (URL: = https://www.sleuthkit.org/autopsy/images/v3/overview.png)

Limitation of Auto Spy

(1) auto spy may experience sluggish performance while analysing extensive datasets, especially if an excessive number of ingest modules are chosen. This can considerably prolong duration of the study.
(2) While autopsy accommodate numerous prevalent file system (such as NTFS, FAT32 and Ext2),it may encounter difficulties with proprietary or obscure file formats, hence constraining its efficacy in certain instances.
(3) Certain complex functionalities necessitate third party modules potentially leading to compatibility or update complications.
(4) While the platform is accessible for fundamental activities advanced functionalities necessitate specialist expertise, and the interface may be daunting for novices.

(2) **Browser History Examiner: -** The Browser History Examiner scrutinizes the web browsing history of the Chrome, Firefox, and Internet Explorer web browsers on the Windows operating system. Browser History Examiner is a forensic software program designed to capture, extract, and analyze internet history from the primary web

browsers installed on a desktop computer. Various forms of data can be evaluated, such as website visits, searches, downloads, and cached files (Fig. 4)

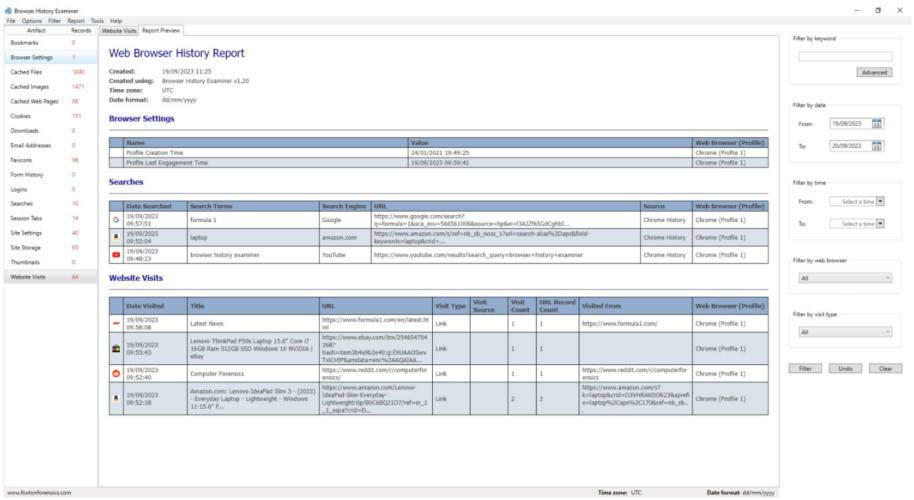


Fig. 4. Browser Examiner forensics (URL = https://www.foxtonforensics.com/browser-history-examiner/)

Limitation of Browser History Examiner

(1) Browser Compatibility: BHE primarily targets major browsers such as Chrome, Firefox, Edge, and Internet Explorer, perhaps restricting its applicability in the examination of less prevalent browsers or mobile platforms.
(2) Forensic Image Support: Although BHE facilitates forensic image analysis, it necessitates prior mounting of the image. This introduces an additional step and necessitates supplementary tools for adequate image mounting.
(3) Data Overload: Similar to numerous forensic tools, the examination of substantial quantities of browser data can be daunting, particularly when addressing significant browsing histories. Analyzing and analyzing this data might be laborious.
(4) Deleted Data Recovery: While BHE can retrieve deleted history via Volume Shadow Copies, this functionality is restricted to Windows systems where such copies exist, which may not consistently occur.

(3) **Net Analysis:** - Net Analysis is a software application created to extract, analyze, and display digital evidence about internet browsing and user behavior on computers and mobile devices. Our Net Analysis suite includes an advanced data recovery tool specifically designed for recovering lost browser artifacts. These objects can be brought into Net Analysis for additional examination. Net Analysis is a cutting-edge software solution that offers substantial improvements over existing programs and methods (Fig. 5).

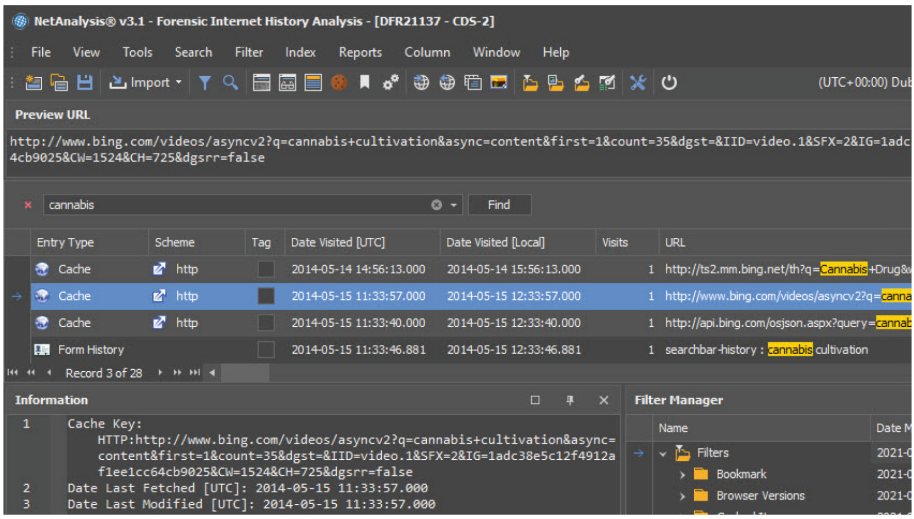


Fig. 5. Net Analysis (URL = "https://www.digital-detective.net/product/netanalysis/")

Limitation of Net-Analysis:-

(1) **Cross-platform Issues**: Although Net Analysis is compatible with multiple operating systems and browsers, the examination of artifacts from obscure or obsolete browser versions may be less effective, necessitating extra work or third-party tools.

(2) **Human Error and Interpretation**:- The precision and efficacy of any forensic tool's analysis predominantly rely on the user's expertise and experience. The potential of misinterpretation or neglecting critical data throughout the analysis process is ever-present.

(3) **Data Recovery Limitations**: Although Net Analysis is compatible with multiple operating systems and browsers, the examination of artifacts from obscure or obsolete browser versions may be less effective, necessitating extra work or third-party tools.

(4) **Internet Evidence Finder**: - The Internet Evidence Finder (IEF) is a software application utilized in the field of computer forensics. Its primary objective is to extract data from various sources such as hard drives, live RAM, or files, with the intended purpose of collecting evidence on the internet. The IEF software was purposefully designed to meet the special requirements of digital forensics examiners. Furthermore, the IEF approach is extensively employed by specialists in the fields of IT information security, electronic discovery, cyber security, and corporate investigations are areas of expertise (Fig. 6).

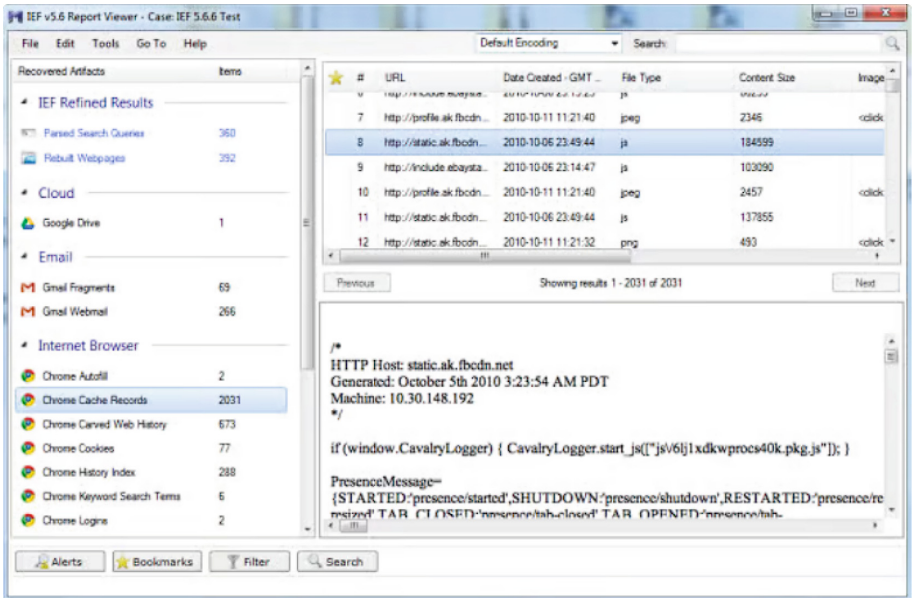


Fig. 6. Internet evidence finder (https://www.forensicfocus.com/reviews/internet-evidence-finder-ief/)

Limitation of Internet evidence finder

(1) Processing Speed: IEF's capacity to extract substantial data from several sources might result in sluggish performance, especially when handling extensive datasets, live RAM, or intricate file systems.
(2) False Positives:- The tool's extensive search techniques may occasionally result in the retrieval of extraneous data or false positives, necessitating additional time for filtering by forensic analysts.
(3) Limited File System Support:- Although IEF accommodates standard file systems, it may encounter difficulties with proprietary or obscure systems, hence constraining its applicability in specific specialized studies.
(4) Training and Expertise:- Despite being intended for forensic examiners, the instrument necessitates a significant learning curve for optimal utilization, and improper application may result in incomplete or overlooked evidence.
(5) Cost:- IEF is a proprietary software, and its licensing costs may be excessive for smaller investigation teams or organizations with financial limitations.

3.2 Identify Evidence Target of the Browser is as Below

See Table 3.

Table 3. Identify evidence target of browser

Tool Name	Target Web Browser	Get the Evidence
Disk Internals Linux Leader (x64)	Chrome, Firefox	Cookies, URLs, Downloads, Cache, Search Keywords
IECacheView 1.5.5.0	Internet Explorer	Cache data (Webcache. Index. Dat)
Index. Dat Viewer 4	Internet Explorer	Index. Dat (history)
Autopsy	Multiple	History, Cookies, Cache, Downloads, Bookmarks, Search History, and more
EnCase	Multiple	Comprehensive browser analysis, including deleted data recovery and advanced reporting
FTK	Multiple	Extensive browser forensics capabilities, including data extraction, analysis, and visualization
Net Analysis	Multiple	Network-related evidence, including browser history and DNS queries
Browser History Examiner	Multiple	User-friendly interface for viewing and analyzing browser history
BrowsingHistoryView	Internet Explorer	Viewing Internet Explorer history

3.3 Keeping a Record in the Computer

Web browsers are used to save user activity in various areas of the operating system. Analyzing data is crucial in many fields for acquiring consumer information. Additionally, the data changes depending on the particular web browser being utilized. User data in web browsers is stored in four separate locations. These items consist of cookies, cache logs, and browsing history logs. Web browsers claim that data is stored in various directories and locations within the system. Examining data from various directories is crucial in the analysis phase. It is essential to seek folders for each of the four designated types of records. Internet Explorer is frequently used by computer users for browsing the web. The browsing history for every user is kept in their profile folder on the Internet. The data is stored individually in the Cookies, Cache, Download History, and history directories as indicated in the Table. The index.dat or container.dat database file holds data stored in directories, containing binary information. The Safari browser stores web data in a specific format within a binary file called "history" found in the browser.

3.4 Clearing Web Browser History

After analyzing the web browser data, it became apparent that there existed a potential for intentional deletion of the browser data. End-users of web browsers are provided with the capability to remove their cache, cookies, browsing history, and downloaded files. The removal of this information will result in the deletion of data that is accessible through a browser function by users. There exist two distinct methodologies for the removal of records. The initial method employed for data loss entails the overwriting of pre-existing data upon the initiation of smartphone applications. The alternative approach entails the user engaging in manual deletion of content from the browser's menu or index. Retrieving historical data using the initial methodology poses significant challenges. The retrieval and erasure of the disk according to the second process enable the retrieval of data (Table 4).

4 Proposed System

Table 4. Enumerates the many tools and pathways available to browser users to delete records.

Web Browser	Delete Options Path
İnternet Explorer	Settings/ Internet Options/ / Deletes
Firefox	Settings /Privacy/History about:preferences#privacy
Google Chrome	Settings /History/Search Data chrome://settings/clearBrowserData
Safari	Settings / Privacy / Delete All Web Site Data Settings/History
Opera	Settings /History/Privacy and Security/Delete All Search Data opera://settings/clearBrowserData

4.1 Methodology for Extraction of Evidence and Reconstruction Data in Browser Forensics

1. **Acquisition of Forensic Image**

   Objective: Obtain a forensically sound image of the suspect's storage device containing browser-related data.

   Method: Utilize trustworthy forensic software to generate an identical copy (forensic image) of the storage device or relevant partitions. To safeguard the integrity of the primary evidence, it is advisable to employ write-blocking hardware or software.

2. **Identification of Targeted Browser Artifacts**
   Objective: Identify and locate specific artifacts within the forensic image that are relevant to the investigation, such as browsing history, bookmarks, cookies, cache files, and download records.
   Method: Utilize forensic software tools to scan known locations where browsers store data, such as SQLite databases, configuration files, and browser cache directories.
3. **Data Extraction and Parsing**
   Objective: Extract and parse relevant data from identified artifacts to retrieve information pertinent to the investigation.
   Method: Employ forensic tools capable of parsing SQLite databases and other relevant file formats. Retrieve URLs, visit frequency data, timestamps, and additional metadata related to your browsing behavior.
4. **Reconstruction of Browsing Activities**
   Objective: In order to understand the suspect's online interactions and behavior, construct a coherent chronology of their browsing patterns.
   Method: Reconstruct the sequence of websites accessed, searches conducted, downloads performed, and interactions on web pages by comparing the obtained timestamps with URLs and other metadata.
5. **Analysis of Extracted Data**
   Objective: Analyze the obtained data to identify relevant trends, correlations, and patterns pertinent to the research.
   Method: To evaluate navigation patterns, consider the frequency of visits to specific websites, search queries, and other user interactions, and use analytical methodologies and data visualization tools. Identify patterns or anomalies that may provide clues to suspect intent or behavior.
6. **Documentation and Reporting**
   Objective: Document the results obtained and the procedures implemented during the extraction process to ensure transparency and accountability.
   Method: Create a comprehensive report detailing the methodology used, tools used, results obtained, interpretations adopted, and challenges faced when collecting and analyzing data related to web browsers. Ensure that the report meets legal standards for admissibility.
7. **Quality Assurance and Validation**
   Objective: We use a comprehensive quality assurance protocol to ensure the reliability and accuracy of the data obtained.
   Method: Cross-reference results with other digital evidence sources (where applicable), verify the consistency and accuracy of extracted timestamps, and perform validation tests to ensure data integrity and reliability.
8. **Legal Compliance and Chain of Custody**
   Objective: For evidence to be admissible in court, it must strictly adhere to a series of administrative to follow rules and regulation and legal requirements.
   Methods: during this process compile every documentation step by step, compile process are including data extraction, processing, data storage, analysis, as well as management, transfer, and storage of forensics data & image. This process is followed as a strictly to protect the dependability and uniformity of the evidence.

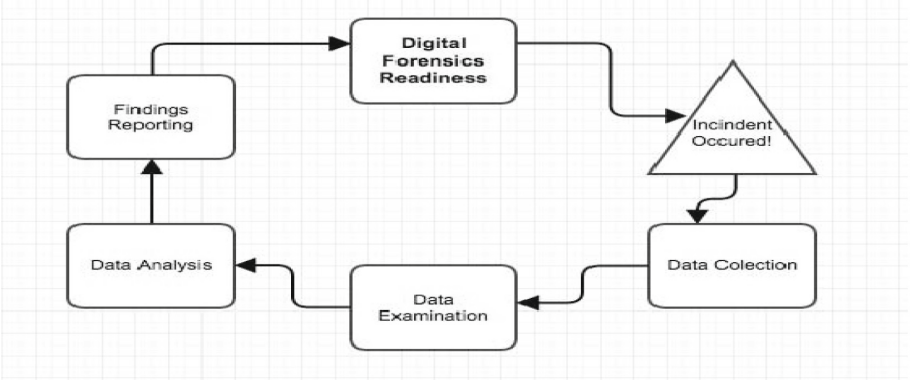


Fig. 7. Process to get evidence in browser forensics (URL: "https://www.researchgate.net/figure/Digital-Forensic-Investigation-Lifecycle_fig1_304037254")

4.2 Types of Algorithm are Used in Browser Forensics

(1) **SQLite Query and Recovery Algorithms:-** most modern browser, like chrome and firefox store their data(like history, cookies bookmarks) in SQLite database. Forensics algorithms focus on querying these database, retriving deleted or lost data using this algorithm in tool like sqlite forensics toolkit or sqolite deleted record parser. These algorithm can recover deleted record using page level analysis and the write ahead logging mechanism.

(2) **Keyword search Algorithms:** this algorithm name is boyer-morre string search algorithm. This is a string searching algorithm used to specific keywords or patterns in browser history, cache or logs. The boyer-moore algorithms is efficient because it skips sections of data that do not match the keyword speeding up the search process. This is essential for searching large browser data file like history database or search logs.

(3) **Hashing Algorithms:**- cryptographic hashing algorithms like MD5 ad SHA-256 are used in browser forensics to verify the integrity of digital evidence, such as history files, cache, and download files. Hash values are computed for each file and can be used to detect any tampering or changes. This ensure the device collected from browser remains authentic.

(4) **Data carving Algorithms:** This algorithm are used to recover deleted or fragmented browser data (ex cache file, download).they scan the raw disk for specific file signatures(like html, jpeg, png etc.) and attempt to reconstruct the file without file system metadata. Tool like foremost and scalpel are commonly used data carving.

(5) **Compression Decompression Algorithms:** browser often store cache data in compressed formats to save space.L277(used in Gzip), Brotil and other compression algorithm are used to compress or decompress cache files.in browser forensics, these algorithms are applied to extract and analyze cached web pages, image and other resources stored in compressed formats.

(6) **Timestamp Conversion Algorithms:**- these algorithms are used to browsers store many of their artifacts(such as cookies, history entries, and log file) with timestamps

in Unix epoch time(the number of seconds since January 1,1970).Algorithms are used to convert these timestamps into human-readable dated to correlate user activities with specific times. Tool like **gmttime()** and strftime() are commonly used for this purpose

(7) **Machine learning for Anomaly Detection:**- in advance browser forensics, machine learning algorithm like random forest and support vector machines (SVM) are used to detect anomalies or unusual behavior in browsing data. The set algorithm can be trained on typical browsing patterns and then used to detect outliers that might indicate malicious activity.

(8) **Timeline Reconstruction Algorithms:**- Time reconstruction in browser forensics involves creating a sequence of events(ex-web visits file downloads) based on artifacts like history, cookies and logs. DAG model user actions as nodes and edges while pathfinding algorithms help trace and understand the sequence of events that occurred browsing sessions. This helps investigation visualize a clear timeline of the user's online activity.

(9) **Geolocation Analysis Algorithms:** IP geolocation algorithms are used to extract location-based data from browser artifacts like cookies, logs or history.by analyzing the user's IP addressed, forensics tool can map approximate geolocation data, which helps in determining the user's location during specific browsing sessions.

4.3 Methodology for the Analysis of Evidence in Browser Forensics

(1) **Timeline Analysis**: digital forensics require precise determination of the subject's movement patterns and meticulous monitoring of the temporal sequence of events. To obtain a comprehensive understanding of the Suspects' unlawful activities, the investigator can do a meticulous chronology analysis. This assessment makes it easier to track how the user navigates across web sites and records how the user interacts with each one individually. Moreover, it is crucial to consider the statistical data about various temporal zones. As indicated in Sect. 3.3, Coordinated Universal Time (UTC) is employed by all five prominent web browsers. Therefore, the time data obtained from the log file may not always match up with the current time of the person being looked into. Therefore, the researcher must utilize a time zone adjustment method for the temporal data. Therefore, The individual conducting the investigation will not be able to establish the specific length of time in which the suspect engaged in online activity in their local area. When conducting an analysis of log data for a suspected occurrence in New York (UTC/GMT), it is imperative for the investigator to appropriately modify the time region of New York to ensure correct examination of the time statistics.

(2) **Analysis Search History:** It is imperative to thoroughly analyze the search terms that the suspect inputted into the search engine, alongside doing an investigation into the websites he has accessed. The aforementioned search terms may also generate keywords about his transgression, which might encompass a solitary word or maybe an entire statement. In this particular case, the use of terms serves as substantiation of the suspect's endeavors to acquire information about his criminal activities, perhaps unveiling the underlying intent, objective, and strategies employed. After doing a search engine query, the selected search phrases are stored as HTTP URL data.

Figure 7 depicts the hierarchical nature of the general HTTP URL information. The Path component of this structure serves as evidence that the appropriate HTTP URL was employed for conducting search operations. In addition, the variable name offers more details regarding the search terms. As an illustration, when an individual inputs the Enter the phrase “cyber security” in the search box of Google, the corresponding Universal Resource Locator (URL) is generated.This HTTP URL provides a wealth of information, including the website, google.com, and the route (Fig. 8),

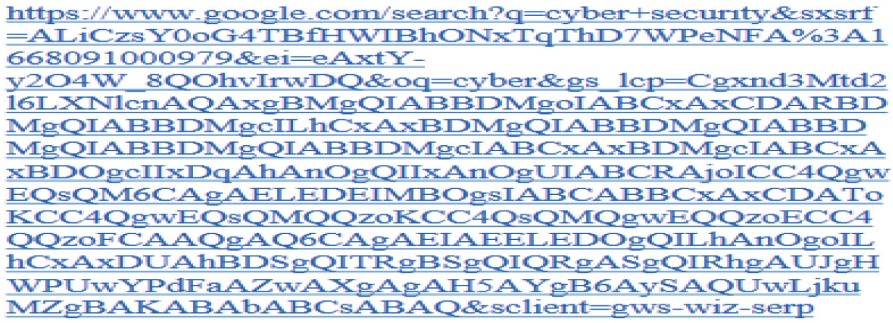


Fig. 8. Analysis search history

The HTTP URL supplied includes a wide variety of details, covering the host domain, google.com, and the designated path, search. This document provides important details about HTTP URLs linked to search functions. After setting the variable q, it becomes clear what search keywords the suspect prefers. In simple terms, the variable q represents the search terms. Different search engines use different language to describe the host, path, and variable parameters. Hence, it is crucial to analyze the HTTP URL structure used by different search engines. After the authors’ examination, it was found that Google, Yahoo, Baidu, Bing, Ask, AOL, Excite, Lycos, Alta Vista, and MSN were the top ten search engines worldwide. Certain assumptions are made when creating URL addresses for HTTP. Primarily, the term "search" is present in the majority of host and path names on various search engines. Most search phrase variables are typically referred to as query, p, or q.

These theoretical concepts aid in identifying search terms from an unknown HTTP URL, as long as the word "search" appears in the host and path names, as well as in variable names q or p. Information retrieval systems that are not familiar with this method need to create a new signature database and take out important search terms from it. By using this method, a detective can determine the search terms used by a suspect, which helps them understand the process, goal, and motive behind the crime.

(3) **Analysis of URL Encoding**:-

The variable q must be found to locate search terms, but decoding the meaning of encoded search terms is not possible using this method. To show encoded characters in HTTP URLs, use hexadecimal codes and the % symbol, which precedes each one-byte character. Each website utilizes a distinct encoding approach (Fig. 9).

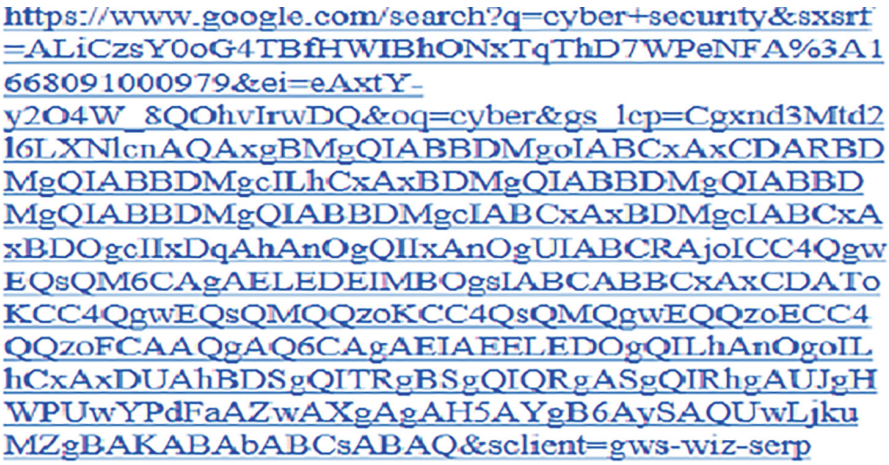


Fig. 9. URL Encoding

The majority of the top 10 search engines globally use UTF-8 encoding on their websites. Baidu primarily utilizes GB2312 encoding, but it also employs Unicode encoding for non-Chinese search queries. In East Asia, search engines generally use DBCS encoding for their operations. There is no standardization in DBCS encoding. Moreover, Japanese search engines like Livedoor make use of Encoding techniques such as EUC-JP and Shift-JIS. As mentioned before, the majority of search engines use encoding techniques from the UTF-8, Unicode, or DBCS categories for search queries and other characters. At times, some search engines use various encoding methods to display multiple words in one HTTP URL Having these distinctive markings. With it being there, the investigator can confirm that Unicode is the encoding being utilized. The investigator may employ the UTF-8 encoding scheme described in RFC 3629 to use a distinct bit pattern for encoding in UTF-8. Encoding in binary using a pair of bytes. Therefore, it is impossible to use DBCS encoding. Recognizable by employing UTF-8 encoding with two bytes.UTF-8 encoding includes a form of two-byte encoding. Therefore, distinguishing between DBCS encoding and two-byte UTF-8 encoding is impossible. An option is to use a search engine-driven plan to distinguish the decoding techniques and display all the identified word strategies. Through the utilization of this methodology, the researcher can distinguish between various encoding techniques and implement a suitable decoding strategy for words. Consequently, this will facilitate comprehension of the encoded words.

(4) **An examination of user behaviour**: A single HTTP URL is inadequate for determining a suspect's online conduct in a sequence of web browser activities during an investigation. Determining the classification of the suspect's activities based on HTTP URL data would facilitate the monitoring of the websites they have accessed and enable the prediction of their previous moves. An evaluation of user behavior requires the researcher to examine all relevant web pages. Employing this approach requires the researcher to actively interact with the Internet browser, which can be a

time-consuming procedure. An algorithm is needed to compute user activity based on HTTP URL data to efficiently expedite the investigation process for digital forensic analysis. The website administrator has the authority to decide the content specified in the HTTP URL. The attributes of an HTTP URL, such as the domain and path, provide the exact address of the webpage. In most cases, an HTTP URL contains a description that clearly defines the intended purpose or function of the webpage. This functionality facilitates the categorization of the user's online behavior by analyzing a certain word found in the HTTP URL address.

4.4 Case Study: The Craigslist Killer-Browser Forensics Investigation

Background

The infamous "Craigslist Killer" case involved Philip Markoff, a medical student accused of using Craigslist to lure victims under the guise of responding to classified ads. Markoff was linked to multiple assaults and the murder of a woman in a Boston hotel. Investigators needed to connect him to the online activities that led to the criminal incidents, particularly his use of Craigslist to contact victims. To achieve this, browser forensics played a crucial role in piecing together his online behavior and providing digital evidence for the case.

Problem

Markoff used several digital methods to conceal his identity, including using multiple devices and attempting to delete incriminating browser history and messages related to the criminal activities. Investigators faced the following challenges:

1. **Recovering Deleted Browser History**: Markoff had attempted to clear his browsing history and delete any trace of interactions with his victims.
2. **Linking Online Activity to Criminal Acts**: Investigators needed to definitively prove that Markoff had used Craigslist to contact the victims and arrange meetings.
3. **Maintaining Data Integrity**: Since browser forensic evidence would be presented in court, it was critical that investigators maintain the integrity of the data to avoid any challenges to its admissibility.

Solution

1. **Recovering Deleted Browser History Using SQLite Recovery**: Investigators began by analyzing Markoff's personal laptop and mobile devices. Most modern web browsers, like Chrome and Firefox, store browsing data in SQLite databases, even after the user clears their history. Using forensic tools like **FTK Imager** and **Autopsy**, the team was able to recover deleted entries from Markoff's browser history databases, specifically URLs and search terms related to Craigslist ads. Specialized algorithms helped reconstruct interactions that had been previously deleted.
2. **Analyzing Craigslist Communications via Cache and Cookies**: Even though Markoff deleted his browser history, the investigators knew that **web cache and cookies** would retain data from his sessions on Craigslist. Using tools like **CacheBack**

and **Browser History Examiner**, they extracted cached files from his browser, which included remnants of web pages, images, and Craigslist sessions. This allowed investigators to reconstruct what had been viewed and which pages he had visited, showing his interactions with specific classified ads related to his victims.

3. **Keyword Searches to Trace Craigslist Messages**: Forensic investigators used **keyword search algorithms** to search for specific terms and usernames that matched the Craigslist ads posted by Markoff or by the victims. **Boyer-Moore string search algorithms** were employed to search through browser artifacts, including cached HTML pages and recovered logs, for key phrases and usernames tied to the criminal acts.
4. **Timeline Reconstruction to Correlate Digital and Physical Events**: Investigators then used **timeline reconstruction algorithms** to establish a sequence of events that correlated Markoff's online activity with the timing of the crimes. By analyzing timestamps from Craigslist searches, emails, and other browser activity, they connected his online communications with the victims to the dates and locations of the assaults and murder. These timestamps also helped investigators establish Markoff's presence in the locations where the crimes occurred.
5. **Cross-Device Correlation and IP Address Tracing**: Markoff used multiple devices to communicate with his victims, including his smartphone and laptop. Investigators analyzed both devices for browser history and Craigslist activity using tools like **X1 Social Discovery**. They also used IP address logs obtained from Craigslist to match Markoff's internet activity to the same IP addresses used in the attacks. This cross-device correlation provided conclusive evidence that Markoff had used these devices to communicate with his victims.
6. **Hashing for Data Integrity**: Throughout the investigation, investigators used **hashing algorithms** such as **SHA-256** to maintain the integrity of the digital evidence. This ensured that the recovered browser data, communications, and other digital artifacts had not been tampered with during the investigation. The hashes were used to authenticate the data in court, providing an unbroken chain of custody.

Outcome

Using browser forensics, investigators were able to recover deleted browsing history, Craigslist messages, and interactions that tied Markoff to his victims. The digital evidence played a key role in linking him to the crimes and ultimately led to his arrest. Despite his efforts to cover his tracks, the detailed forensic investigation allowed law enforcement to present a clear, methodical timeline of his activities, connecting his online behavior to the criminal acts.

Conclusions

In summary, browser forensics offers a systematic and reliable approach to collecting, securing and analysing digital evidence related to online browsing activities. Investigator can quickly locate and examine key artifacts like browsing history, cookies, cache and form data to ensure the accuracy and sufficiency of the evidence.by integrating automated techniques with human verification, the process becomes more reliable while reducing the risk of data manipulation or errors. Traditionally, theses risks were carried out using separate tools, each handling specific aspects of the comprehensive. However, researcher

will develop comprehensive frameworks that will consolidate all these functions into a single, streamlined process. This unified approach will only enhance the efficiency and accuracy of browser forensics investigations but will also significantly advance the filed of digital forensics by enabling more thorough and precise analyses of online activities.

References

1. Akbal, E., Güneş, F., Akbal, A.: Digital forensic analyses of web browser records. J. Softw. **11**(7), 631–637 (2016). https://doi.org/10.17706/jsw.11.7.631-637
2. Divyesh, G., Dharan, D., Nagoor Meeran, A.R.: Forensic evidence collection by reconstruction of artifacts in portable web browser. Int. J. Comput. Appl., 32–33 (2014a)
3. Divyesh, G., Dharan, D., Nagoor Meeran, A.R.: Forensic evidence collection by reconstruction of artifacts in portable web browser. Int. J. Comput. Appl. **93**, 32–33 (2014b)
4. Kang, J., Lee, S., Lee, H.: A digital forensic framework for automated user activity reconstruction. In: ISPEC 2013, vol. 7863–7863, pp. 263–277 (2013)
5. Kauser, S., Malik, T.S., Hasan, M.H., Akhir, E.P., Kazmi, S.M.H.: Windows 10's browser forensic analysis for tracing P2P networks' anonymous attacks. Comput. Mater. Continua (Print) **72**(1), 1251–1273 (2022a). https://doi.org/10.32604/cmc.2022.022475
6. Kauser, S., Malik, T.S., Hasan, M.H., Akhir, E.P., Kazmi, S.M.H.: Windows 10's browser forensic analysis for tracing P2P networks' anonymous attacks. Comput. Mater. Continua (Print) **72**(1), 1251–1273 (2022b). https://doi.org/10.32604/cmc.2022.022475
7. Majeti, G., Yvl, S.S., Ulichi, S.S., Mohanty, S.N., Sudha, S.V.: Digital forensic advanced evidence collection and analysis of web browser activity. ICST Trans. Scal. Inf. Syst. (2023). https://doi.org/10.4108/eetsis.3357
8. Mehta, S.D., Upadhyay, D.: A review on classification of Tor-Nontor traffic and forensic analysis of tor browser. Int. J. Eng. Res. Technol. (IJERT) **9**(04), 776 (2020). http://www.ijert.org
9. Mohamed, S.M., Abdelbaki, N., Shosha, A.F.: Digital forensic analysis of web-browser based attacks. In: International Conference Security and Management, p. 237. CSREA Press (2016). https://URL_TO_PDF
10. Nalawade, A., Bharne, S., Mane, V.: Forensic analysis and evidence collection for web browser activity. In: Ramrao Adik Inst. of Technology, p. 518. International Institute of Information Technology (I^2IT), Pune (2016). https://doi.org/10.1109/ICACDOT.2016.7536506
11. Oh, J., Lee, S., Lee, S.: Advanced evidence collection and analysis of web browser activity. Digit. Investig. **8**, S62–S70 (2011). https://doi.org/10.1016/j.diin.2011.05.008
12. Rathod, D.M.: Web browser forensics: google chrome. Int. J. Adv. Res. Comput. Sci. **8**(7), 896–899 (2017). https://doi.org/10.26483/ijarcs.v8i7.4433
13. Singh, A., Singh, S.K., Nayak, S.K., Singh, N.: Solid state technology cyber-crime and digital forensic: challenges resolution. Solid State Technol., 166–167 (2021). https://www.researchgate.net/publication/351245795
14. Singh, A., Singh, S.K., Vege, H.K., Singh, N.: A framework for crime detection and diminution in digital forensics (CD3F). Int. J. Adv. Comput. Sci. Appl. **13**(9), 332–333 (2022). https://www.ijacsa.thesai.org
15. Umar, R., Yudhana, A., Faiz, M.N.: Experimental analysis of web browser sessions using live forensics. Int. J. Electr. Comput. Eng. (IJECE) **8**(5), 2951–2958 (2018). https://doi.org/10.11591/ijece.v8i5.pp.2951-2958
16. Marrington, A., Baggili, I., Al Ismail, T., Al Kaf, A.: Portable Web Browser Forensics. Zayed University, College of Information Technology (2021)

DDOS Attack Detection and Prevention Using MIEC Method in SDN Environment

R. Karthikeyani(✉) and E. Karthikeyan

PG and Research Department of Computer Science, Government Arts College, Udumalpet, Tamilnadu, India
keyanirangu@gmail.com

Abstract. The Software Defined Network (SDN) is one of the best commonly used network management architecture in recent years. The main aim of SDN is enabling flexibility and simplicity in network operation and management via centralized controller. This approach is allowing administration to efficiently oversee, configure, and optimize various network devices like switches, routers, and access points from a unified control point. SDN encounters numerous challenges, including latency issues, scalability, packet forwarding, security and cost. Among this security is a major concern for SDN, particularly due to the threat of Distributed Denial of Services (DDOS) attacks, which pose a significant risk to the SDN environment. Despite ongoing advancement in tools and technology, the detecting and mitigating DDOS attack within SDN remains a challenging task. As of now no fully efficient solution has been developed to effectively cope with the challenges posed by DDOS attack in SDN environment. The proposed Mutual Information Entropy Calculation method (MIEC) is a rapid and simple solution for detecting DDOS attack and improve previous entropy detection algorithm. Experimental result further shows more accurate outcomes and faster reporting in comparison to previous methodologies.

Keywords: DDOS · centralized controller · SDN · mutual information · entropy

1 Introduction

Nowadays, exponential growth of technology in last few years has led to an increasing the number of network devices, highlighting the growing importance of network service control. Demand for high-level control has grown as well due to the increasing the number of network devices exceeding earlier specification [1, 2]. The traditional networks face a lots of problems. Such as problem with scalability, configuration and management and flexibility. The main goal of SDN is to use of centralized controller to get in addition to the drawback of traditional network [3]. SDN's ability to separate the control plane from the data plane is the one of its primary objective [4]. In traditional networks, routing decisions are taken by the routers itself [5, 6]. However, in other hand SDN controller taking the routing decisions which develop network management. [6]. In Fig. 1 illustrates the primary differences between SDN network architectures and traditional network architecture and this figure emphasize the advantages of SDN and it specially focusing

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 165–175, 2025.
https://doi.org/10.1007/978-3-031-86305-9_12

centralized control and improved flexibility compared to traditional architecture. SDN have a three-tier architecture comprising of i. Application layer, ii. Control layer, and iii. Infrastructure layer [7].

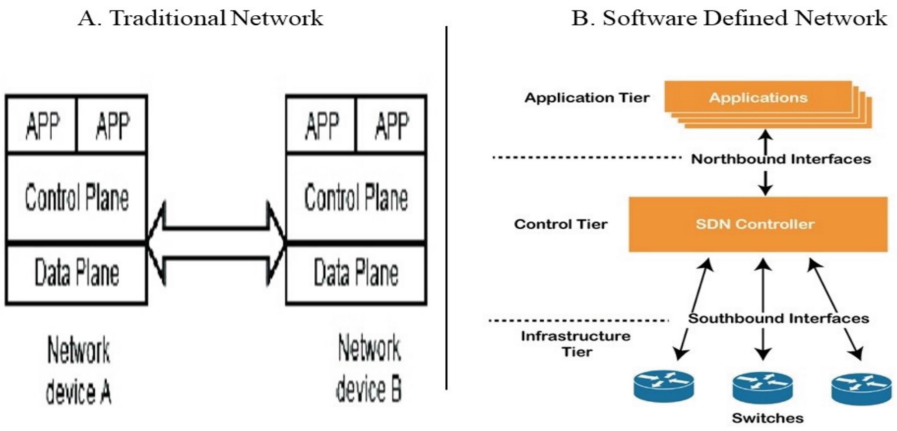


Fig. 1. Different between Traditional Network and SDN

- **Application layer** contain application and function for the network. This layer connected to the control layer use of north API interface. This layer allows to administrator for program control functions in network [8].
- **Control layer** act as an intermediary between the application layer and management layer. In SDN controller play main role in the all network control operations in the control layer. The controller handle all forwarding packets at the lower layer [8, 9].
- **Infrastructure layer** is responsible for performs essential functions support communication and data transfer. It include hardware devices, such as switches, routers and storage devices for network that perform packet forwarding functions and communicate controller via open flow protocol [8, 10].

Centralized controller has many advantages including simplified network management compared to the traditional network [11]. However, this centralized control also poses challenges and weaknesses, particularly concerning security [12]. If the SDN controller is compromised or attacked, it can have serious implications for the entire network. One of the most common and severe threat is DDOS attacks, where malicious actors flood the controller with a massive volume of traffic, overwhelming its process capacities and disrupting network operations [13, 14]. DDOS attacks can cripple the functionality of the SDN controller, leading to network downtime, service disruptions, and potential data breaches. Proposed method implements robust security measures to protect the control layer from such attacks, including intrusion detection systems, access controls, and traffic monitoring and anomaly detection mechanisms [15]. Furthermore, proactive

monitoring, regular security audits, and timely software updates are essential to mitigate the risks associated with SDN controller vulnerabilities. Successful management of these security concerns is vital [16].

The proposed method in this context utilizes MIEC for attack detection, with a focus on detecting DDOS attacks. This approach is considered one of the simplest and most straightforward methods for DDOS attack detection. MIEC method is employed to gauge the level of information shared between two random variables. This measurement indicates the degree of dependence or correlation between these variables. In this context of attack detection, MIEC method helps in identifying potential DDOS attack. By leveraging MIEC, the proposed method aims to enhance the detection capabilities for DDOS attacks, providing a reliable and efficient means of identifying and mitigating such threats in network environment.

The paper is organized as follows Sect. 1 clearly explains Introduction of SDN. Section 2 explains related work. Section 3 explains the proposed detection mechanism. Section 4 explains the Experiment Environment and Result discussion. Section 5 concluded and future work of proposed method. Section 6 References.

2 Related Works

Dinh Thi Thai Mai et al. [8] proposed a method Dynamic Entropy Calculation method, which calculate the entropy value dynamically. This method offers the advantage of being lightweight and very simple approach. It adjusts the threshold value based on the characteristics of input data. Here author build a new dynamic threshold value that changes over time. This method has some demerit also it harder to interpret and understand the inner working of the algorithm.

Cong Fan et al. [12] introduced method utilizing a novel entropy calculation technique called Fusion entropy. This approach combines the log energy entropy and information entropy to enhance the detection of DDOS attacks. The method is noted for its rapid detection capabilities and offers a flexible and robust framework. However, a notable drawback of this method is the challenge of determining appropriate weights for various elements or categories.

Anupama Mishra et al. [18] proposed a method based on Shannon Entropy calculation, which is noted for simplest and light weight in nature. The method employs a pox controller for DDOS attack detection and utilized Mininet for network simulation and Wireshark for network analyzer. The simulation involves 64 hosts and a single controller. However, a significant limitation of this approach is its reliance on static threshold value for DDOS detection coupled with the use of single controller. These factors may restrict the method's effectiveness and scalability in dynamic and large-scale network environments.

Jisa David et al. [2] proposed a method using fast entropy approach for DDOS attack detection. This method involves the rapid analysis of network traffic with entropy values calculated continuously for each flow of request. This will reduce a computational time. Here the threshold value is updated adaptively based on traffic characteristics. This improves detection speed. This method primarily focuses on statistical anomalies traffic based on entropy changes. Setting and maintaining appropriate entropy threshold

requires ongoing monitoring and tuning based on network attack patterns. Failure to adjust threshold may lead to missed detections or false alarms.

Lying Li et al. [6] proposed a method to calculate entropy. This approach leverages the randomness of packets headers and introduces a novel concept known as Cumulative Entropy Calculation. While the method recognized for its simplicity and effectiveness in detecting DDoS attacks, several drawbacks need to be addressed. This paper does not thoroughly explore the scalability implications of their entropy-based algorithm. In large scale network environment with high traffic volumes, the continuous calculation of entropy could lead to significant computational overhead, potentially affecting the methods efficiency and real-time effectiveness.

Vitalii Savchenko et al. [15] proposed a method detecting slow DDOS attacks through time delay forecasting method. It identifies subtle, for low-rate DDOS attacks that often go traditional detection methods. By leveraging time delay forecasting techniques, this paper demonstrates the ability to detect anomalies in network traffic patterns indicative of so DDOS attacks. It enhances against such stealthy threats. Its potential drawback is complexity in implementation, and sensitivity to network conditions, dependence on historical data.

Ye J.et al. [20] proposed a method using Support Vector Machine in SDN, an effective machine learning approach. It suitable for detecting various types of DDOS attacks with reasonable accuracy. It provides timely alerts and enabling rapid response mechanism. But this method has some demerits. It is complex and high dimensional feature spaces. it requires significant computational resource and memory.

3 Proposed Detection Mechanism

In this section, we will initially outline the structure of the defensive mechanism we propose, subsequently, provide a comprehensive overview of the training process and algorithm utilized for detection purpose.

3.1 Entropy Based Detection Approach

Entropy is a fundamental concept used in various field such as information theory, physics, and statistics. It represents the measure of uncertainty, randomness, or disorder in a system or dataset [17, 18]. The concept originated in thermodynamics, where it was used to describe the tendency of energy in a closed system to disperse and reach equilibrium. There are many types of entropy depending on the context in which it is used. They are Shannon's entropy, Renyi joint entropy, Boltzmann entropy, Information entropy entropy, Relative entropy, and Conditional entropy [18–20]. Entropy calculation can be used as a method to detect abnormal or malicious network traffic patterns [21, 22]. The formulas for Shannon's entropy, a fundamental concept in information theory, is given have as follows:

$$H(X) = -\sum_i P(x_i).\log_2 P(x_i) \tag{1}$$

where, $H(X)$ denotes entropy of given dataset X. This measures the average uncertainty or information content within the data. $P(x_i)$ represent the probability of occurrence of each distinct value [23, 24].

The entropy calculation involves summing the products of each probability of $P(x_i)$ and its corresponding logarithm base2, with a negative sign applied to ensure a non-negative entropy value. This formula is widely used across various domains, machine learning, and network security, to quantify randomness, uncertainty, or information content in data sets and probability distributions.

In the context of DDOS attack detection, entropy proves to be efficient method for identifying abnormal or malicious network traffic patterns. Building on this concept, proposed method a novel approach known as MIEC. This new entropy-based method aims to enhance DDOS attack detection by utilizing the concept of mutual information, which measures the amount of information obtained about one variable through another.

3.2 The Architecture of Proposed Work

Entropy values are calculated based on window size and threshold values. A window size means amounts of data send or received by a system. The proposed method incoming packets are continuously monitored, and their mutual entropy values are assessed. If multiple packets from different sources arrive at the same port of a host or switch within a predefined window of time and surpass a predetermined threshold value, it indicates a potential for DDOS attack. The controller maintains a log of the calculated mutual entropy of each packet. If the entropy value remains consistently above the set threshold within a predefined time window during an attack, it signals a substantial ongoing attack. Upon detecting an attack targeting a specific switch is then blocked to prevent additional malicious traffic from infiltering the network.

Flowchart of proposed method is given in Fig. 2. This architecture first collects the flow of incoming packets then check if it has reached window size 50. Once 50 packets are reached, mutual information entropy is calculated. In this experiment we used a POX controller and an openflow switch. If the calculated entropy value exceeds the threshold, DDOS attack is detected, and the corresponding port number ID blocked. Algorithm 1 explains the proposed method. In this proposed method, focusing on detecting flood based DDOS attacks by generating packets using scapy. Scapy is an exceptionally potent tool capable of tasks such as scanning, sniffing, attacking, and generating packets.

Here Mutual Information entropy is a measure of the amount of information that is between two random variables. It quantifies the reduction in uncertainty about one variable when the value of the other variable is known. The formula for mutual information entropy is derived from information theory and is expressed as:

$$I(X;Y) = \sum_{x\epsilon X} \sum_{y\epsilon Y} P(x,y)\log\left(\frac{P(x,y)}{P(x),P(y)}\right) \tag{2}$$

Here, $P(x, y)$ is the joint probability of X and Y taking values x and y simultaniously, while $P(x)$ and $P(y)$ are their individual probabilities

Algorithm for DDOS attack detection using MIEC method

1. Let Probability distribution of X and Y: P(x) and P(y).
2. Joint probability distribution of X and Y: P (x, y).
3. Compute the entropy of X: $H(X) = -\sum_{i=1}^{n} P(x_i) \log_2(P(x_i))$.
4. Compute the entropy of Y: $H(X) = -\sum_{j=1}^{m} P(y_j) \log_2(P(y_j))$.
5. Calculate the joint entropy of X and Y: $-\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 P(x_i,\ y_j)$.
6. Calculation the mutual information between X and Y:
   $I(X;Y) = H(X) + H(Y) - H(X,Y)$.
   This equation quantifies how much information X and Y share or how much knowing on reduces uncertainty about the other. (Mutual information I (X; Y) represent the amount of information shared between X and Y.)
7. if (MIEC>Threshold) then
8. generate the alert message of DDOS attack detected
9. block port no
10. else
11. forward the packet
12. end if

4 Experimental Environment

This proposed experiment was worked operating system Linux ubuntu 22.04. Using Mininet emulator 2.3.0. Network topology created using Mininet. 64 host and 9 switches on one controller using for experimental structure. This proposed system when turn on the controller that controller observe the traffic in the network during simulation.

4.1 Pox Controller

In this proposed system, it is essential to choose the appropriate controller. At present many open-source controllers are available. For this experiment, we selected pox controller. It is open source and lightweight and very easiest controller. The pox controller, developed by Stanford University, is an integral component of SDN and it is written primarily in the Python programming language. It will be based on OpenFlow protocol. Pox has the simplest architecture and provides excellent performance for researchers. For these reasons, we selected the pox controller for the simulation experiment.

4.2 Result Discussion

In this proposed method experiment each test involves sending 250 data packets with a window size of 50 packets (amount of data sent at a time). Each test involves sending packets 5 times with a 10 s interval. For every window size the entropy value calculated. In normal scenario host 2 and 3 send randomly normal traffic to network topology. In attack scenario host 1, host 2, host 4, host 5, host 6 send attack traffic to target host 16. Compare the two scenarios this method the minimum entropy 0.5 was adapted as the threshold of our experience. Table 1 shows our proposed method entropy values normal and attacking scenarios.

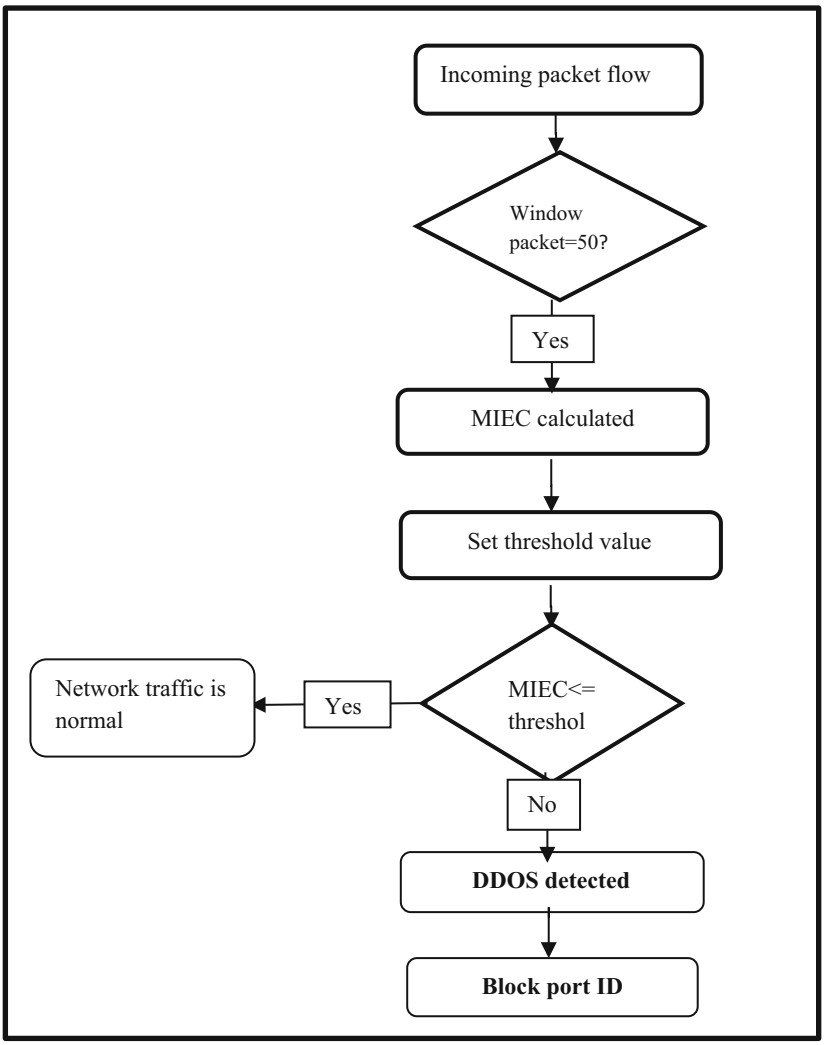


Fig. 2. Flow chart of proposed method

In Fig. 3 illustrates the comparison of proposed method entropy variations in two scenarios. Among the 35 packets in the graph one plot represent the normal scenario MIEC entropy variations. Another plot represents Attacking scenario MIEC entropy variation. A minimum entropy of threshold 0.5 is used in the proposed method. The entropy value increases gradually with every 5^{th} packet. In the attacking scenario, the highest recorded entropy value is1.90. When the entropy value is exceeds the threshold, the corresponding port will be blocked.

Using same simulation set up without changing existing entropy calculation method used to measure the randomness of network traffic. Table 2 shows three different information of existing method and proposed method. I. How much time taken for reaching the

Table 1. MIEC method entropy value normal scenario and attacking scenario

Incoming packets	Normal scenario	Attacking scenario
5	0.07	0.71
10	0.32	0.90
15	0.34	1.04
20	0.40	1.09
25	0.41	1.15
30	0.47	1.36
35	0.49	1.90

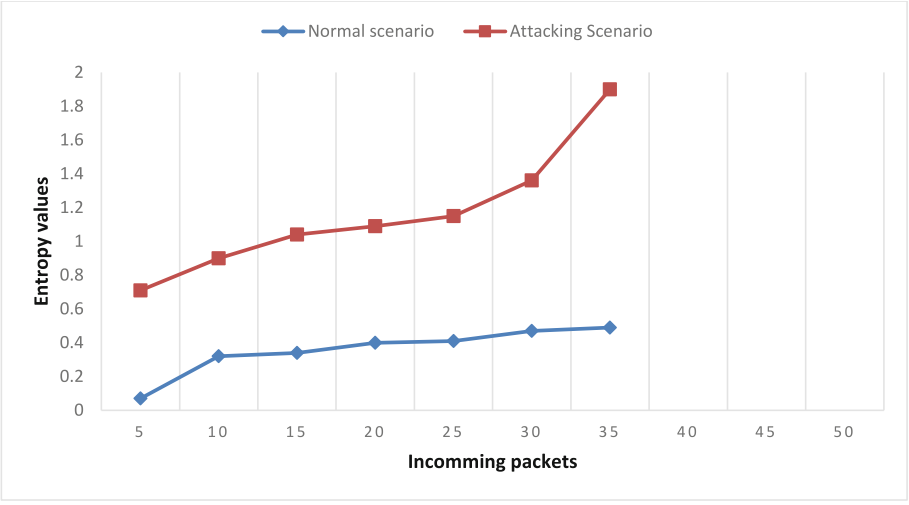


Fig. 3. Graph for MIEC method entropy value normal scenario and attacking scenario

adopted threshold value 0.5. II. How many packets needed for attack detection III. Existing and proposed system entropy values. In Fig. 4 clearly explained proposed method is best and fastest.

The proposed method offers several significant advantages over existing methods for detecting DDOS attacks:

a. **Reduction time:** Existing methods requires approximately 30 s detecting a DDOS attack. In contrast, the proposed system reduces this detection time to just 13 s, enhancing the responsiveness of the attack detection process.
b. **Fewer packets required for detection:** Existing method detects a DDOS attack after receiving 65 packets from the same IP address. The proposed method, however, is able to detect an attack with only 55 packets, making it more efficient in identifying malicious activity with fewer data points.

Table 2. Comparison between existing method and proposed method

No of packets	Existing system entropy	Time for existing method	No of packets proposed method	Proposed system entropy	Time For proposed method
10	0.11	6	5	0.07	2
25	0.27	13	10	0.32	3
32	0.31	15	15	0.34	5
45	0.34	19	20	0.40	10
50	0.37	21	25	0.41	12
60	0.43	25	30	0.47	17
65	0.47	30	55	0.49	25

c. **Fewer host required:** While existing systems requires monitoring from 6 hosts to efficiently detect a DDOS attack, the proposed system can achieve the same detection capabilities with only 3 hosts. This reduction not only simplifies the deployment but also decrees the computational and resource requirement.

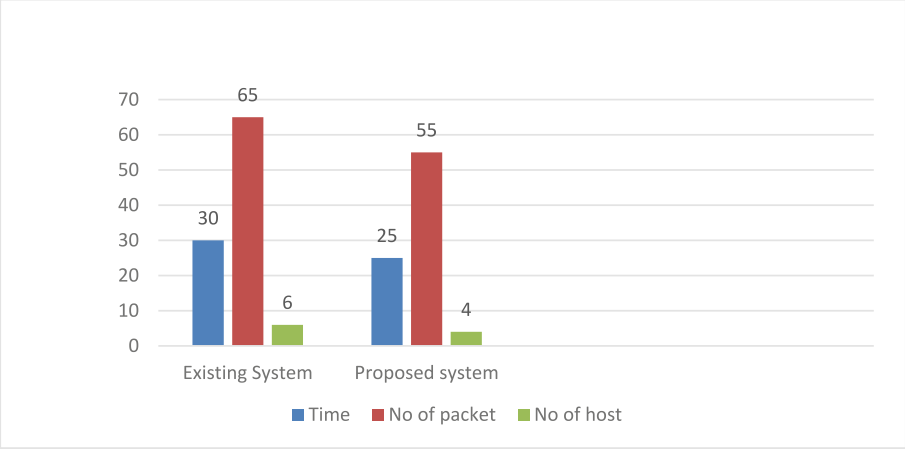


Fig. 4. Comparison between existing system and proposed system

5 Conclusion and Future Work

This paper mainly focuses on detecting DDOs attacks targeting SDN controllers. The proposed method uses the MIEC method, which is an advanced approach compared to traditional entropy calculation method. This method more effective in detecting DDOS

attack compared to previous normal entropy calculation methods. Additionally, the proposed method enables rapid detection by the controller. In future, DDOS attack detection might involve using multiple controllers, which introduces a single point of failure and may make data recovery impossible if the controller fails. However, using multiple controller s overcome this issue and enhances the system resilience.

References

1. Koay, A., Chen, A., Welch, I., Seah, W.K.G.: A new multi classifier system using entropy-based features in DDoS attack detection. In: 2018 International Conference on Information Networking (ICOIN), Chiang Mai, Thailand, pp. 162–167 (2018). https://doi.org/10.1109/ICOIN.2018.8343104
2. David, J., Thomas, C.: DDoS attack detection using fast entropy approach on flow-based network traffic. Procedia Comput. Sci. **50**, 30–36 (2015). https://doi.org/10.1016/j.procs.2015.04.007
3. Kalkan, K., Gür, G., Alagöz, F.: SDNScore: a statistical defense mechanism against DDoS attacks in SDN environment. In: 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, Greece, pp. 669–675 (2017). https://doi.org/10.1109/ISCC.2017.8024605
4. Yang, Zhao, H.: DDoS attack identification and defense using SDN based on machine learning method. In: 2018 15th International Symposium on Pervasive Systems, Algorithms and Networks (I-SPAN), Yichang, China, pp. 174–178 (2018). https://doi.org/10.1109/I-SPAN.2018.00036
5. Isong, B., Molose, R.R.S., Abu-Mahfouz, A.M., Dladlu, N.: Comprehensive review of SDN controller placement strategies. IEEE Access **8**, 170070–170092 (2020). https://doi.org/10.1109/ACCESS.2020.3023974
6. Li, L., Zhou, J., Xiao, N.: DDoS attack detection algorithms based on entropy computing. In: Qing, S., Imai, H., Wang, G. (eds.) ICICS 2007. LNCS, vol. 4861, pp. 452–466. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77048-0_35
7. Ujjan, R.M.A., Pervez, Z., Dahal, K., Khan, W.A., Khattak, A.M., Hayat, B.: Entropy based features distribution for anti-DDoS model in SDN. Sustainability **13**, 1522 (2021). https://doi.org/10.3390/su13031522
8. Mai, D.T.T., Dat, N.T., Bao, P.M., Truong, C.Q., Tung, N.T.: DDOS attacks detection using dynamic entropy in software-defined network practical environment. Int. J. Comput. Netw. Commun. (IJCNC) **15**(3) (2023). https://doi.org/10.5121/ijcnc.2023.15207
9. Dayal, N., Srivastava, S.: Analyzing behavior of DDoS attacks to identify DDoS detection features in SDN. In: 2017 9th International Conference on Communication Systems and Networks (COMSNETS), Bengaluru, India, pp. 274–281 (2017). https://doi.org/10.1109/COMSNETS.2017.7945387
10. Dong, S., Abbas, K., Jain, R.: A Survey on distributed denial of service (DDoS) attacks in SDN and cloud computing environments. IEEE Access **7**, 80813–80828 (2019). https://doi.org/10.1109/ACCESS.2019.2922196
11. Yan, Q., Yu, F.R., Gong, Q., Li, J.: Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: a survey, some research issues, and challenges. IEEE Commun. Surv. Tutor. **18**(1), 602–622 (2016). https://doi.org/10.1109/COMST.2015.2487361
12. Fan, C., Kaliyamurthy, N.M., Chen, S., Jiang, H., Zhou, Y., Campbell, C.: Detection of DDOS attacks in software defined networking using entropy. Appl. Sci. **12**(1), 370 (2022). https://doi.org/10.3390/app12010370

13. Aladaileh, M.A., et al.: Dynamic threshold generalized entropies-based DDOS detection against software-defined network controller. Appl. Sci. **13**, 775 (2023). https://doi.org/10.3390/app13020775L
14. Xu, Y., Liu, Y.: DDoS attack detection under SDN context. In: IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, pp. 1–9 (2016). https://doi.org/10.1109/INFOCOM.2016.7524500
15. Savchenko, V., et al.: Detection of slow DDoS attacks based on time delay forecasting. In: III International Scientific and Practical Conference "Information Security And Information Technologies", Odesa, Ukraine, 13–19 September 2021
16. Varghese, J.E., Muniyal, B.: An efficient IDS framework for DDoS attacks in SDN environment. IEEE Access **9**, 69680–69699 (2021). https://doi.org/10.1109/ACCESS.2021.3078065
17. Tan, L., Pan, Y., Wu, J., Zhou, J., Jiang, H., Deng, Y.: A new framework for DDoS attack detection and defense in SDN environment. IEEE Access **8**, 161908–161919 (2020). https://doi.org/10.1109/ACCESS.2020.3021435
18. Mishra, A., Gupta, N., Gupta, B.B.: Defense mechanisms against DDoS attack based on entropy in SDN-cloud using POX controller. Telecommun. Syst. **77**, 47–62 (2021). https://doi.org/10.1007/s11235-020-00747-w
19. Kalkan, K., Altay, L., Gür, G., Alagöz, F.: JESS: joint entropy-based DDoS defense scheme in SDN. IEEE J. Sel. Areas Commun. **36**(10), 2358–2372 (2018). https://doi.org/10.1109/JSAC.2018.2869997
20. Ye, J., Cheng, X., Zhu, J., Feng, L., Song, L.: A DDoS attack detection method based on SVM in software defined network. Secur. Commun. Netw., 1–8 (2018). https://doi.org/10.1155/2018/9804061
21. Aladaileh, M., Anbar1, M., Hasbullah, I.H., Sanjalawe, Y.K., Chong, Y.-W.: Entropy-based approach to detect DDoS attacks on software defined networking controller. Comput. Mater. Continua. https://doi.org/10.32604/cmc.2021.017972P
22. Tian, Q., Miyata, S.: A "DDoS attack detection method using conditional entropy based on SDN" traffic. IOI **4**, 95–111 (2023). https://doi.org/10.3390/iot4020006
23. Mousavi, S.M., St-Hilaire, M.: Early detection of DDoS attacks against SDN controllers. In: 2015 International Conference on Computing, Networking and Communications (ICNC), Garden Grove, CA, USA, pp. 77–81 (2015). https://doi.org/10.1109/ICCNC.2015.7069319
24. Yu, S., Zhang, J., Liu, J., Zhang, X., Li, Y., Xu, T.: A cooperative DDoS attack detection scheme based on entropy and ensemble learning in SDN. EURASIP J. Wirel. Commun. Netw. **2021**(1), 1–21 (2021). https://doi.org/10.1186/s13638-021-01957-9

A Lossless Image, Video and Audio Encryption Method Based Hybrid Chaotic Map and DNA Encryption

Vrushali Khaladkar[1,2(✉)] and Manish Kumar[2]

[1] Department of Mathematics, Fergusson College (Autonomous), Pune, Maharashtra, India
p20200511@hyderabad.bits-pilani.ac.in

[2] Department of Mathematics, Birla Institute of Technology and Science, Pilani, Hyderabad Campus, Jawahar Nagar, Kapra Mandal, Medchal District, Hyderabad 500078, Telangana, India

Abstract. The suggested lightweight video encryption strategy offers a promising method for safeguarding multimedia data in cyberspace. The proposed strategy is based on a hybrid chaotic map and DNA encryption. Due to the fewer dimensions of the chaotic map, it can be adapted to secure real-time data generated using closed-circuit television and other Wireless sensors network tools. The strategy provides a strong level of security and resistance against different attacks by utilizing advanced cryptographic techniques including hybrid chaotic systems and DNA encryption. The proposed work is decently compared with other contemporary competing encryption schemes which shows that the proposed one can be adapted for the encryption of videos in lightweight applications. The thorough assessment using simulation results and statistical experiments highlights the practical applicability and efficiency of the proposed encryption technique.

Keywords: Encryption · Decryption · DNA encryption · Hybrid chaotic modified logistic map

1 Introduction

In the era of big data, along with data science and analytics, data security plays a very important role. Security of data is mainly considered under two categories viz. during transmission and during storage [13].

Security during transmission needs encryption schemes to be light weight and with less time complexity. The data under consideration is in various formats like textual data, image data, voice data and video data which is transmitted and stored in the form of bits which is accessible to everyone and hence vulnerability

The authors are very grateful for providing the necessary facilities and motivation to carry out this research.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 176–195, 2025.
https://doi.org/10.1007/978-3-031-86305-9_13

increases towards various attacks. Hence during transmission as well as during storage data security is of utmost priority and if a hacker tries to hack the information, then the information must be in ambiguous form so that guessing is also difficult.

Cryptography and cryptology are a science of securing information which comes under applied Mathematics. Technique of transforming important information into unreadable format is called encryption and the reverse process is referred to as decryption. For daily routine transactions we unknowingly share our crucial information on social platforms like use of ATM cards, use of biometrics for various processes and use of identity cards based on RFID-radio frequency identification. These transactions share text, image data routinely. The authentication of such information can be implemented using cryptographic algorithms.

The strength of such algorithms can be tested using various statistical tests along with the robustness against various attacks. The major two categories of such algorithms involve symmetric key algorithm and public-private key algorithm. The encryption process is mainly encompassed using confusion and diffusion processes. Classical algorithms like RSA, ELGamal, Transposition, Hill cipher, LFSR etc. are used for confusion and diffusion processes. The proposed encryption scheme is based on 1-dimensional hybrid chaotic map along with DNA encryption as a pivotal part.

2 Literature Review

The validation of stability and efficiency of any mathematical encryption algorithm against various attacks and the strength through statistical analysis is a key aspect. Several researchers have worked on Chaotic nature of dynamical systems and their use in encryption schemes [1,2,4,7,8,11,14,17,19,20,23]. Fusion of discrete dynamical systems and fractals is given by the research [1]. The work [2] explains the protection of multimedia data using various chaotic maps. The author's [1] study proposes an n-dimensional chaotic model for resolving the shortcomings of existing chaotic systems. Multimedia type data is secured using various algorithms like digital signatures and water markings, information hiding like steganographic algorithms is explained by the authors [3,4]. Multiple alterations along with chaotic dynamical systems have been designed for information security [1,8,13–16]. Fast, new key generation algorithm is proposed by the authors [4,5] based on 1-D logistic map for securing medical images while the author [4] has presented watermarking based encryption scheme for securing medical images. Cellular Automata is used to generate chaotic sequence which eventually used as encryption key presented by researchers [9]. An efficient optical image encryption system based on double random phase encoding and Arnold's Cat map is presented by researchers [10]. In the research work [6], researchers have proposed a novel multi-key encryption technique for improving security in video telecommunication based on the algorithms viz. RSA and elliptic curve cryptography. DNA encryption along with high dimensional chaotic

maps are used for strong security and how the algorithm works robustly against various attacks is given by [8,21]. The work proposed by authors in [11] shows how chaotic neural networks in synchronization with each other are used for image encryption. Researchers [18] shows how efficiently 2-D hybrid Henon maps are used for generating chaotic keys for encryption schemes for the multimedia type data and audio generated from voice messages. The work given by [22] gives an image cryptosystem for data transfer in cloud-based smart cities using the cascading of Logistic-Chebyshev and Logistic-Sine maps.

In the proposed research, hybrid chaotic map 1-D MLM along with DNA encryption is used for encryption of separated audio and RGB-frames. The confusion is carried out using DNA encryption while diffusion using 1-D MLM chaotic map. The proposed color video encryption algorithm is strongly recommended for day-to-day life data transactions where light weight security is required. Due to low dimensional chaotic maps, and high security, this algorithm works better for biometric data security, voice security at critical audios and videos. Encryption algorithms like data encryption standards, advanced encryption standards, elliptic curves show high complexity. Security analysis is tested and performed on various video types but only a subset of videos is presented here. The strength of the proposed algorithm is tested using statistical analysis and its effectiveness to handle attacks is tested using various methodologies.

3 Description of the Proposed Methodology

The RGB video encryption scheme proposed here consists of following steps: Separation of audio from video, separation of frames from video, scrambling using DNA encryption scheme, confusion using chaotic map.

3.1 Chaotic Map

This section shows how 1-D MLM shows strong chaotic nature with good range of chaotic parameters, large scale of chaotic range, large Lyapunov exponent, etc. along with DNA sequence based encryption scheme. Since the dimension of proposed chaotic map is low so can be strongly considered for the applications where the resources (transmission through like wireless sensors networks) are under constraints.

3.2 1-Dimensional Modified Logistic Map (1-D MLM)

The proposed scheme uses recently proposed 1-D MLM to conquer issues associated with the 1-dimensional chaotic logistic map. Mathematically, 1-D MLM is expressed by Eq. 1:

$$\tilde{O}_{j+1} = \left[\frac{(2^{\mathscr{K}} (\rho \times \tilde{O}_j (1 - \tilde{O}_j)))}{(\sin(\tilde{O}_j))^{\mathscr{R}}} \right] \pmod 1. \tag{1}$$

The initial seed value for generating the chaotic sequence 2 is represented by $\tilde{O}_0$ while the value $\tilde{O}_{j+1}$ of the sequence is recursively obtained by using $\tilde{O}_j$. The other parameters used in this sequence generation are $\rho \in (0, 10)$, the exponents $\mathscr{R} \in [2, 26]$ and $\mathscr{K} \in [1, 3]$, respectively. These parameters are chosen in accordance with the bifurcation diagram in which the chaotic window size is large and hence provides large chaotic range than standard 1-D logistic map.

3.3 DNA Encryption

DNA double helix contains four nuclei-acid bases in a sequential format namely, $A(adenine)$, $C(cytosine)$, $G(guanine)$, and $T(thymine)$, where the complementary pairs are A and T and G and C. The corresponding representation in binary number system is defined by the rule as 00 and 11 are complementary, 01 and 10 are also complementary. The proposed algorithm use C,A, T and G to denote 00, 01, 10, and 11, respectively. If the Gray image is 8-bit image then to represent each pixel, 8-bits are required, and hence 4-DNA symbols are sufficient. If $(x, y)^{th}$ pixel is having intensity k, then its binary representation is $b_1b_2b_3b_4b_5b_6b_7b_8$. This representation is then converted to DNA sequence by above mentioned rule. First level encryption is obtained as follows:

1. DNA key $\alpha_1\alpha_2\alpha_3\alpha_4$ of length 4 is chosen randomly.
2. Numerical equivalent of this key is obtained in matrix form by copying this key in the row and column dimensions same as that of original frame size.
3. The output of above step is resized to create the Key_1.

4 Proposed Encryption Algorithm

The proposed encryption algorithm is designed for encryption of video. The video under consideration has 30 frames per second along with an audio. Each frame of the video is 8-bit 3-D image in RGB or CMY colour scheme. The following steps provided below explains the working of the proposed scheme.

1: Take input as a video.
2: Separate audio data and frames from the video.
3: A frame array is transformed to a row vector R_V.
4: Key_1 is generated using DNA encryption for level 1 encryption of audio and video using the rules explained in Sect. 3.3.
5: The key Key_2 for second level encryption is generated using 1-D MLM. For this seed values as explained in Sect. 3.1 and Sect. 3.2 are considered with the help of level 1 encryption.
6: The generated key Key_2 is resized to the size of image/frame/size of an audio.
7: Final decryption is obtained by $XORing$ the Key_2 with first level encryption.

4.1 Proposed Decryption Algorithm

Decryption of the video is obtained using the above stated steps in the reversed order to obtain the original image from the cipher image. During this process, $(Key_1)^{-1}$ and $(Key_2)^{-1}$ are used to recover the frames and audio and hence video.

5 Flowchart of Proposed RGB Video Encryption Encryption

A flowchart along with the pixel flow of the proposed RGB video encryption encryption algorithm is demonstrated in Fig. 1 and 2.

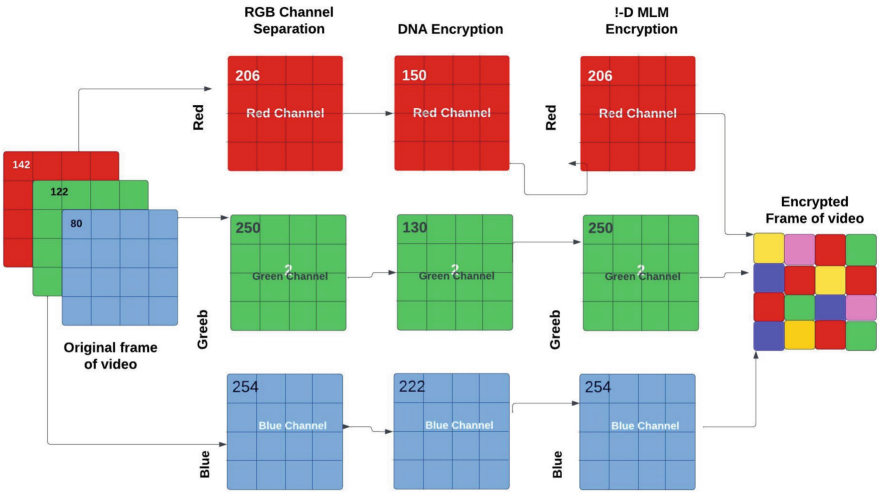


Fig. 1. Proposed encryption flowchart for RGB frames.

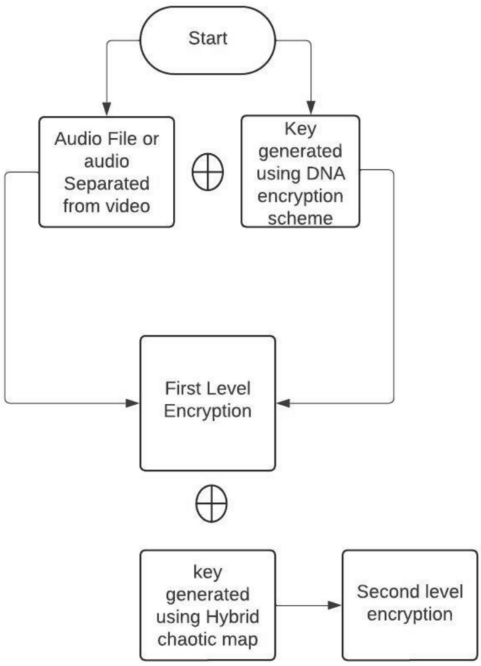


Fig. 2. Proposed encryption flowchart for Audio.

6 Encryption Results

Simulation outcomes of the proposed RGB video encryption strategy are figured out in following Figs. 3 and 4.

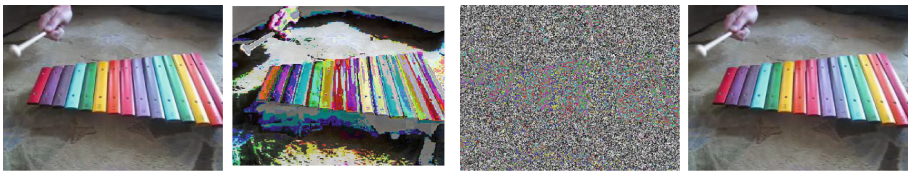


Fig. 3. Simulations results of encryption and decryption of video-1 (a) Plain $frame_1$, (b) First level encryption of $frame_1$, (c) Second level encryption of $frame_1$, (d) retrieved $frame_1$.

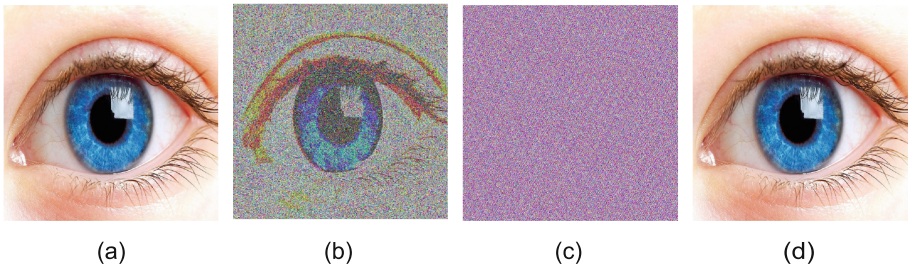


Fig. 4. Simulations results of encryption and decryption of video-2 (a) Plain $frame_1$, (b) First level encryption of $frame_1$, (c) Second level encryption of $frame_1$, (d) retrieved $frame_1$

Encryption results of audio data. These results are tabulated in the Fig. 5

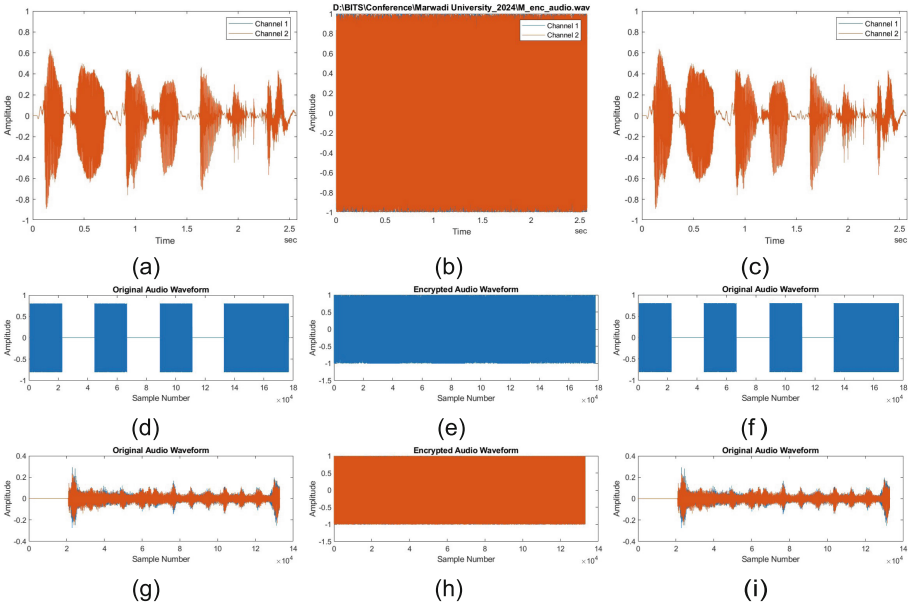


Fig. 5. Encryption results: (a) Original audio, (b) Encrypted audio, (c) Decrypted audio.

7 Performance Analysis of Proposed (1-D MLM)

Through multiple experiments, the chaotic nature and dynamical characteristics of the proposed 1-D MLM are tested. These tests include bifurcation diagram, Lyapunov exponent, phase diagram etc.

7.1 Assessment Using Bifurcation Diagram

Bifurcation diagram of any dynamical system is a graphical representation that describes its stability and chaotic nature. This works well for analyzing the responses of the system to parameter changes. The plot of relationship between the response of dynamical system and parameter changes displays various patterns, transitions in phases and chaotic nature of the system. The bifurcation diagram of the chaotic map used for key generation is given in the following Fig. 6

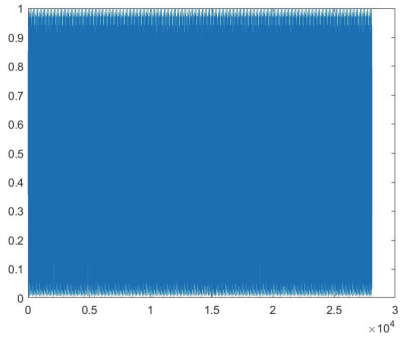


Fig. 6. Bifurcation diagram of 1-D MLM.

7.2 Assessment Using Lyapunov Exponents

To understand sensitivity with respect to the variations in the initial conditions of the dynamical system and the predictability of the chaotic nature in the specific window, the strongest tool is to obtain Lyapunov exponents. Mathematically, Lyapunov exponent is considered as the mean logarithmic rate of segregation or convergence of two neighboring coordinates of two time series $\mathscr{X}_t$ and $\mathscr{Y}_t$ separated by an commencement distance $\delta\rho_0 = \|\mathscr{X}_0 - \mathscr{Y}_0\|_2$. The formula below 2 shows the computation of Lyapunov exponents (λ).

$$\lambda = \lim_{x\to\infty} \frac{1}{n} \sum_{i=1}^{n} ln \frac{|\delta\rho_i|}{|\delta\rho_0|} \tag{2}$$

The dynamical system is said to be chaotic in nature if at least one $\lambda \geq 0$ quantifying the sensitivity to initial conditions.

8 Performance Assessment Through Experimental Results

The authenticity and effectiveness of any encryption scheme is based on the results achieved by carrying various experimental results. These experiments involve diverse statistical tests like histograms, information entropy, correlation, mean square error etc. The strength of the algorithm is verified using different intensity frames from video, also frames which are contaminated using noises and their retrieval rate. The recovery of all frames of the video after cropping specific percentage of one channel of the all the frames also boosts the efficiency of the algorithm. The tabulated results shows the successful execution of the proposed video encryption algorithm for its use in real life applications under lightweight encryption.

8.1 Assessment Using Histograms for Video Frames and Audio

The uniform distribution of intensities in the cipher image confirms the strength of encryption algorithm. The proposed research shows the analysis and calculations of histograms of channel separated frames for before and after encryption. The resultant graphs shows substantial difference. Graphical analysis of various channels is tabulated in Figs. 7 and 8 along with histograms of original audio and encrypted audio.

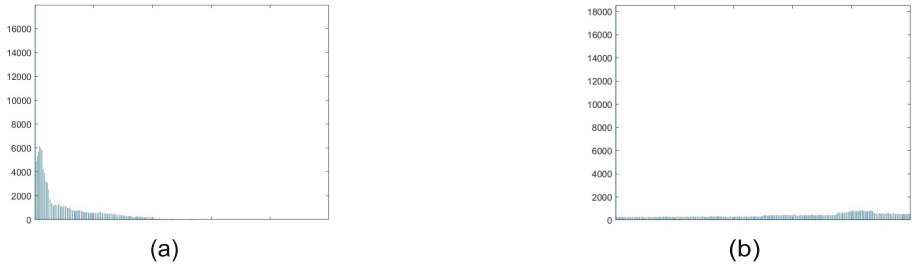


Fig. 7. Analysis through histogram: (Audio of first video) (a) Original, (b) Encrypted.

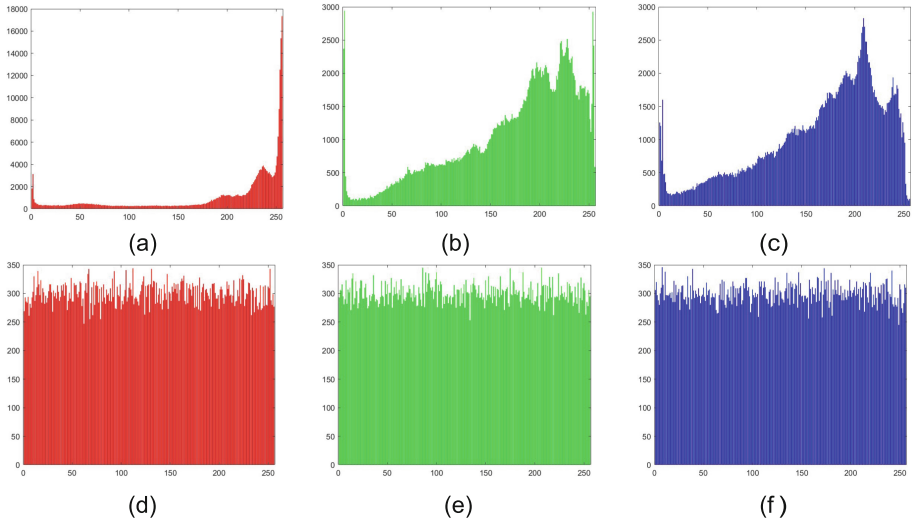


Fig. 8. Assessment using Histograms: (frame1 of video 1) channel wise (a) R plain, (b) G plain, (c) B plain, (d) R cipher (e) G cipher, (f) B cipher.

8.2 Assessment Using Information Entropy

To measure the unpredictability of video or audio data, the most commonly used statistical tool is an information entropy. It is also used to measure the leakage of information during encryption-decryption process. For a video frame, which uses intensity values from $\left[0, 2^{\mathscr{L}} - 1\right]$, an information entropy is approximately $\mathscr{L}$. In this research, the videos of 8-bit frames are used, hence the simulation shows approximate value of information entropy is 8. Mathematically, the image entropy is calculated as Eq. 3:

$$\text{Information-Entropy} = \sum_{i=0}^{2^{\mathscr{L}}-1} \mathscr{P}(\mathscr{M}_i) log_2 \frac{1}{\mathscr{P}(\mathscr{M}_i)}, \tag{3}$$

where,

1. i = Value in the range 0 to $2^{\mathscr{L}} - 1$
2. $\mathscr{P}(m_i)$ = The probability of occurrence of m_i
3. m_i = An event such that the pixel intensity is i

Significantly random video frame has similar probability of occurrence of specific intensities and hence shows information entropy as $\mathscr{L}$. The proposed encryption strategy is defined for 8-bit RGB video, due to the randomness after encryption, the information entropy values are approximately 8. The following Table 1 shows information entropy of various images along with the reference values.

Table 1. Information entropy of $Frame_{11}$ and $Frame_{21}$ channel wise and compared with reference values for the images Baboon and Lena of 512×512.

frame	R-plane	R-plane	B-plane
$Frame_1$	7.9991	7.9973	7.9993
$Frame_2$	7.9995	7.9990	7.9970

(a) **Table 1.**

frame	Proposed	[19]	[20]	[21]
Baboon	7.9991	7.9912	7.9989	7.9969
Lena	7.9997	7.9909	7.9972	7.9970

(b) **Table 2.**

Information entropy is also calculated for the encrypted audio using proposed encryption algorithm. Significantly random audio has similar probability of occurrence of specific intensities and hence shows information entropy as $\mathscr{L}$ where $\mathscr{L}$ represents the number of bits required to store the audio data. The following Table 2 shows information entropy of various audios along with the reference values.

Table 2. Entropy of encrypted Audios

Audio	Entropy
Audio 1	15.4276
Audio 2	15.4216

8.3 Assessment Using Mean Squared Error

To estimate the adjacency of the regression line from the data points is obtained using mean squared error $\mathscr{M}.\mathscr{S}.\mathscr{E}.$. It evaluates the distance from the set of coordinates to the regression line. The errors in this process are the calculated distances and squaring these values gives non-negative real numbers. Mathematically, $\mathscr{M}.\mathscr{S}.\mathscr{E}.$ between the 8-bit RGB frame of the video and the corresponding encrypted frame of size $\mathscr{P} \times \mathscr{Q}$, is given by the equation (4):

$$IMAGE_{\mathscr{M}.\mathscr{S}.\mathscr{E}.} = \frac{1}{\mathscr{P} \times \mathscr{Q}} \sum_{q=1}^{\mathscr{Q}} \sum_{p=1}^{\mathscr{P}} (ENC(p,q) - ORI(p,q))^2, \tag{4}$$

where,

1. $\mathscr{P}$ = Number of rows in plain or cipher frames
2. $\mathscr{Q}$ = Number of columns in plain or cipher frames
3. $ENC(u,v)$ = pixel intensity at the $(u,v)^{th}$ position of the 3-plane encrypted frame,
4. $ORI(x,y)$ = pixel intensity at $(u,v)^{th}$ position of the plain 8-bit 3 RGB/CMY frame.

The mean squared error between the plain 8-bit 3 RGB/CMY frame and cipher 8-bit 3 RGB/CMY frame is given in the Table 7c (Table 3).

Table 3. Channel wise mean squared error between frames before and after encryption and compared with images Baboon, Lena, Peppers of size 512×512, along with MSE between original and encrypted audios.

Frame	R-plane	R-plane	B-plane
$Frame_1$	12067	11906	11713
$Frame_2$	10713	11700	12513

(a) **Table 1.**

Image	Proposed	[22]
Baboon	9060	8900
Peppers	11181	10128
Lena	9074	8948

(b) **Table 2.**

Audio	Proposed
Audio 1	0.4631
Audio 2	0.4626

(c) **Table 3.**

8.4 Assessment Using Correlation Coefficient

For the encrypted frame of video, correlation between any two adjacent pixels in the positive X, positive Y and in diagonal direction is very low. A perfectly secure video/audio should have correlation approximately equal to zero. The tabulated values shows the correlation in positive X, positive Y and diagonal directions for randomly selected 3000 pixels from the frame. The correlation coefficients are determined using these formulas 5–9 and simulation results are tabulated in the Table 7 (Table 4).

$$\hat{\rho}_{pq} = \frac{\text{Covariance(p,q)}}{\sqrt{\hat{\Delta}(p)}\sqrt{\hat{\Delta}(q)}}, \tag{5}$$

$$\hat{\Delta}(p) = \frac{1}{A}\sum_{j=1}^{A}(p_j - \eta(p)^2), \tag{6}$$

$$\hat{\Delta}(q) = \frac{1}{B}\sum_{j=1}^{B}(q_j - \eta(q)^2), \tag{7}$$

$$\eta(p) = \frac{1}{A}\sum_{j=1}^{A} p_j \tag{8}$$

$$\eta(q) = \frac{1}{B}\sum_{j=1}^{B} q_j \tag{9}$$

Table 4. Correlation coefficient between frames and separated audios before and after encryption of varied sizes.

Frame	R-plane	R-plane	B-plane
$Frame_1$	0.0009	0.0011	0.0024
$Frame_2$	0.0014	0.0020	0.0007

(a) **Table i.**

Frame	R-plane
$Audio_1$	-0.0246
$Audio_2$	-0.0091

(b) **Table ii.**

$Frame_1$	R-plane	R-plane	B-plane
Horizontal	0.0042	-0.0348	0.0024
Vertical	-0.0235	0.0018	0.0125
Diagonal	0.0130	-0.0508	0.0090

(c) **Table iii.**

Correlation scatter plots for adjunct pixels in positive Y-axis, positive X-axis and diagonal for the plain and cipher frames are given in the following tabulated Figs. 9 and 10

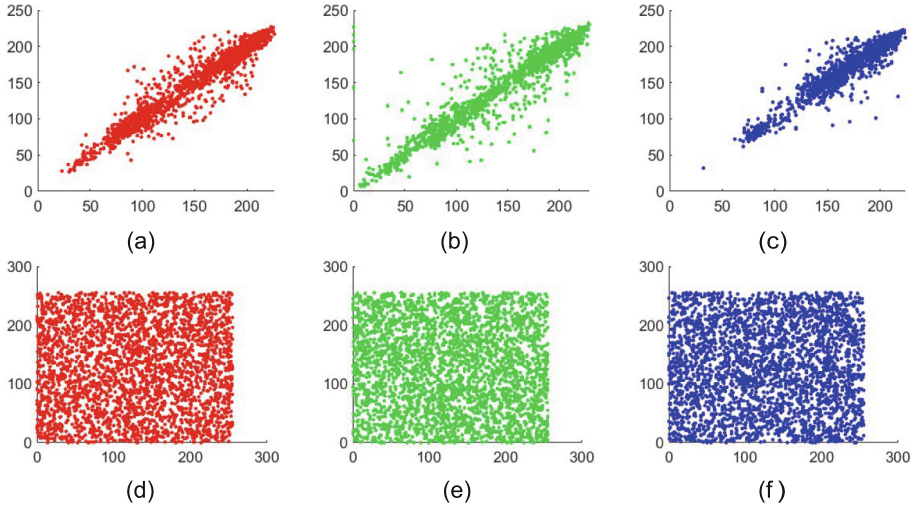


Fig. 9. Analysis through scatter plots: (first frame of video 1 Plain and cipher) (a) R plain, (b) G plain, (c) B plain, (d) R cipher, (e) G cipher, (f) B cipher.

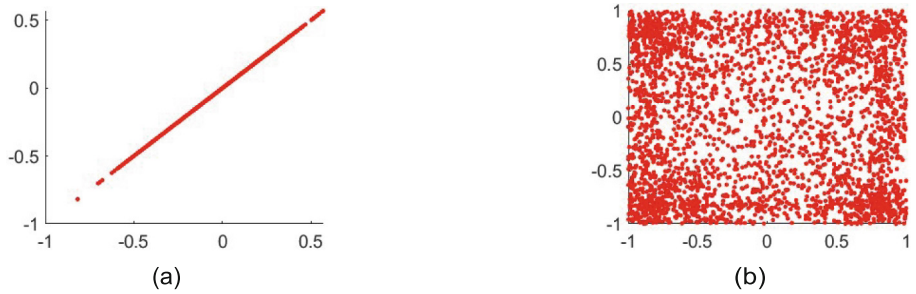


Fig. 10. Analysis through scatter plots: (Audio of first video) (a) Original, (b) Encrypted.

8.5 Peak Signal-to-Noise Ratio

One of the quality measure of audio encryption algorithm is testing the ratio between peak to signal noise. Higher the PSNR-SNR values, better is the encryption scheme and reconstruction of the audio post decryption. Mathematically, PSNR is given by the formula:

$$PSNR = 20 \times Log_{10} \frac{\mathscr{I}_{max}}{\sqrt{\text{Mean square error}}}, \tag{10}$$

where $\mathscr{I}_{max}$ is the maximum available amplitude in the plain audio file. Tables 5 show the PSNR-SNR values for various audio files.

Table 5. PSNR-SNR values for the encrypted audio files with reference to the original audio file.

Audio	Decrypted PSNR	Encrypted PSNR	Decrypted SNR	Encrypted SNR
$Audio_1$	Infinite	6.7504	Infinite	45.6151
$Audio_2$	Infinite	6.0547	Infinite	43.6100
$Audio_3$	Infinite	9.1250	Infinite	35.6151

8.6 Intensity Profile

Image intensity profile is statistical test which checks chaotic nature of the encrypted image and displays the uniformity in image intensity post encryption. The plot includes 3-D representation of *Red*, *Green* and *Blue* intensities and their distribution in original frame and encrypted frame. The plot of original frame shows the dominant intensity is red over green and blue intensities while in encrypted frame, all the three primary intensities are uniformly distributed. The following Fig. 11 displays the intensity profile results.

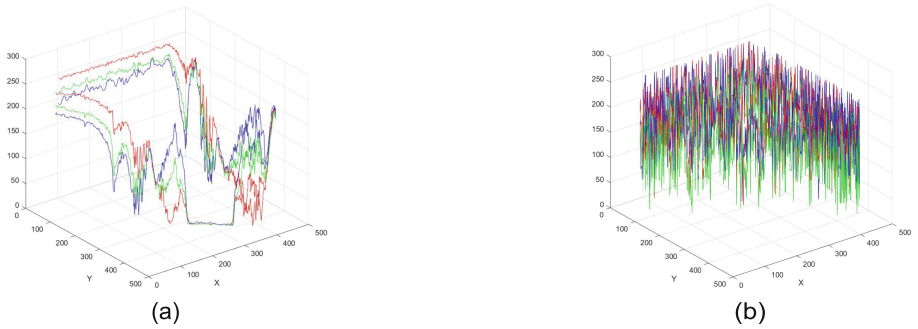


Fig. 11. Intensity profile (a) Plain $frame_1$, (b) Encrypted frame

9 Assessment Using Sensitivity

Sensitivity to the seed values for the dynamical system means each point in the chaotic system is approximated by the other points which have remarkably different trajectories. i.e. an arbitrarily minute variation or perturbation of the original trajectory may lead to significantly different trajectory. This property of chaotic dynamical systems is called butterfly effect. The more the chaotic effect better is the encryption algorithm and hence more the sensitivity, better is the encryption scheme. Sensitivity of the chaotic map is verified using plotting auto correlation, cross correlation graphs with original seed value against the chaotic map with very minute change in seed value. The figures below 12 and 13 shows the auto correlation and cross correlation graphs.

Fig. 12. Analysis through auto correlation: (1-D MLM) (a) Seed value = 8.0 (b) Seed value = 8.000000001

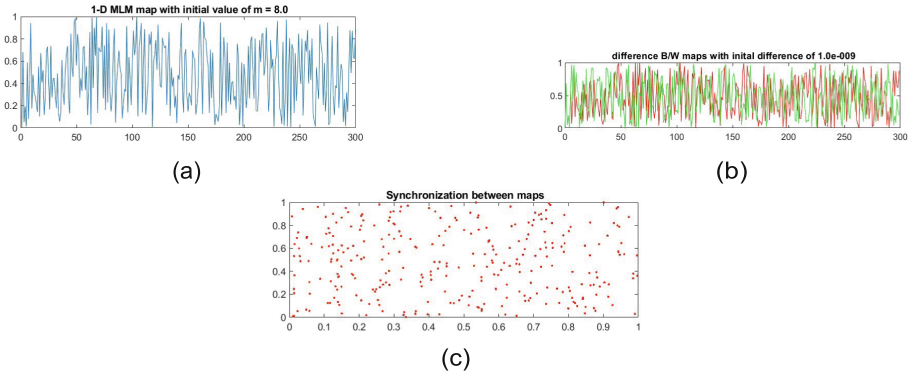


Fig. 13. Analysis through cross correlation: (1-D MLM) (a) chaotic nature with seed value = 8.0 (b) overlaping of chaotic map with seed value 8.0 and seed value = 8.000000001 (c) synchronisation points between two chaotic maps

10 Resistance Against Differential Attack

The metrics used to evaluate effectiveness and efficiency of any encryption strategy is based on its robustness against differential attacks. The study of analysing the differences and alterations percentage due to encryption, is referred as differential attacks analysis. The strongest tools to understand robustness of the

encryption system are NPCR i.e. the Number of Pixels Change Rate (NPCR) or the Number of Bits Change Rate (NBCR) gives the percentage of pixels/bits that are different in the two selected frames and UACI unified average changing intensity. The mathematical representation of the terms (NPCR/NPBR) and (UACI) is given by the following Eqs. 11 and 12

$$NPCR = \frac{\sum_1^B \sum_1^A Distance(P,Q)}{A \times B} \times 100, \tag{11}$$

and

$$UACI = \frac{\sum_1^B \sum_1^A \frac{\mathscr{C}_1(p,q) - \mathscr{C}_2(p,q)}{255}}{A \times B} \times 100, \tag{12}$$

where,

1. A = rows in the frame
2. B = columns in the frame
3. $Distance(P,Q)$ = city block distance between pixels P and Q in the frame
4. $\mathscr{C}_1(p,q) = (p,q)^{th}$ location of the cipher frame encrypted using proposed video encryption strategy.
5. $\mathscr{C}_2(p,q) = (p,q)^{th}$ location of the cipher frame C_2 encrypted using proposed video encryption strategy by making alteration in few pixels or pixel intensities in the original frame.

Table 6. ***npcr/nbcr*** and ***uaci*** results

Three plane frame	*npcr*	*uaci*
$Frame_1$	99.6091	33.4630
$Fame_2$	99.6103	33.4600

The results given in Table 6 indicates the efficiency of proposed video encryption strategy against differential attacks.

11 Assessment Using Noise Injection and Trimming Plane Attack

Adding noise in the encrypted frame and recovering it using proposed decryption algorithm and testing the recovery rate is called adding noise attack analysis. Various noises like impulse, Gaussian noise are added with different percentages and the recovered frames are tabulated below. From 8-bit RGB/CMY frame, any one or two or all the three plane of the frame are cut in two various percentages and these cut frames are recovered using the proposed video encryption strategy. The recovery rate of these frames is tabulated below Figs. 14 and 15

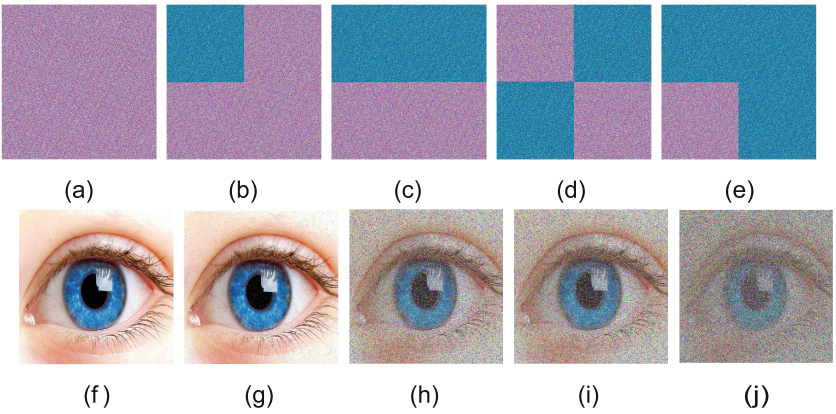


Fig. 14. Trimming channel attack Frame 1 of video 2: (a) Cipher, (b) 25 % trimmed, (c) 50 % trimmed, (d) 50 % trimmed, (e) 75 % trimmed, (f) Retrieved, (g) 25 % Retrieved, (h) 50 % Retrieved, (i) 50 % Retrieved, (j) 75 % Retrieved

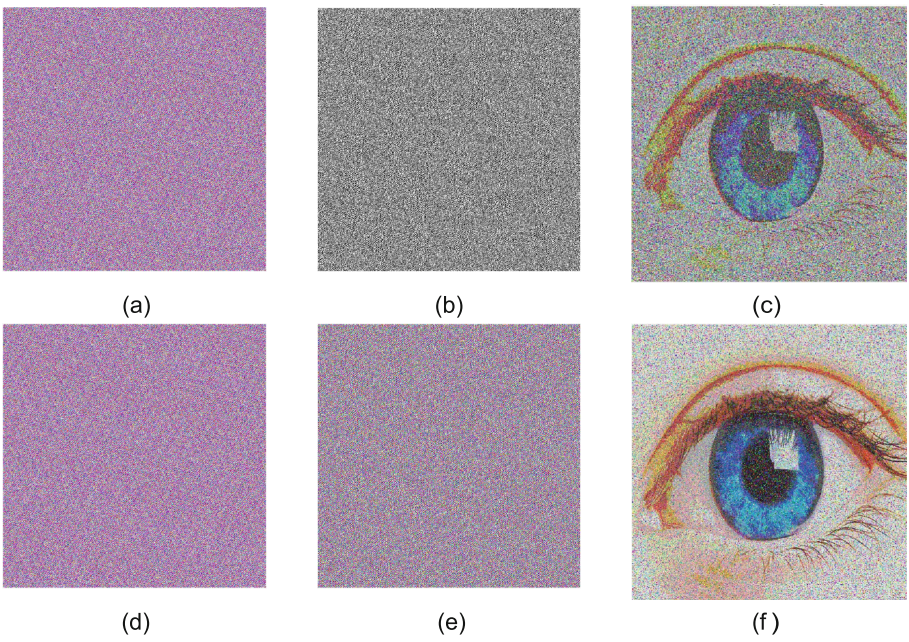


Fig. 15. Injection of noise attacks: (Frame 1 video 2) (a) Cipher frame, (b) injection of 20% Gaussian noise, (c) Retrieved frame, (d) Cipher frame, (e) injection of 20% impulse noise, (f) Retrieved frame.

The recovered images are also tested under various statistical tools for verifying the robustness of the proposed encryption algorithm against the attacks like trimming plane attack and noise injection attack. The results are tabulated below.

Table 7. Statistical analysis of recovered frames from the trimming plane attack.

Frame	correlation
25%*cropped*	0.9585
50%*cropped*	0.6948
75%*cropped*	0.4000

(a) **correlation1.**

Frame	Entropy
25%*cropped*	7.0810
50%*cropped*	7.6245
75%*cropped*	7.4817

(b) **Entropy.**

Frame	Entropy
25%*cropped*	443.2802
50%*cropped*	4052
75%*cropped*	7456

(c) **MSE.**

12 Conclusion

The suggested lightweight video encryption strategy offers a promising method for safeguarding multimedia data in cyberspace. The proposed encryption strategy is based on hybrid chaotic maps and DNA encryption. 1-D modified logistic map (1-D MLM) proposed along with DNA encryption overcomes the issues associated with standard 1-D logistic map, like window size is less, blank and stable windows and chaotic range of 1-D standard logistic map is small. Window size and chaotic range of 1-D MLM is larger than logistic map in spite of having less dimensions. Due to the fewer dimensions of the chaotic map, it can be adapted to secure real-time data generated using closed-circuit television and other WSN tools. The strategy provides a strong level of security and resistance against different attacks by utilizing advanced cryptographic techniques including hybrid chaotic systems and DNA encryption. The proposed work is decently compared with other contemporary competing encryption schemes which shows that the proposed one can be adapted for the encryption of videos in lightweight applications. The thorough assessment using simulation results and statistical experiments highlights the practical applicability and efficiency of the proposed encryption technique.

References

1. Cao, W., Cai, H., Hua, Z.: n-Dimensional Chaotic Map with application in secure communication. Chaos Solit. Fract. **163**, 112519 (2022)
2. Hosny, K.M. (ed.): Multimedia Security Using Chaotic Maps: Principles and Methodologies. SCI, vol. 884. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-38700-6
3. Tuncer, T.: The implementation of chaos-based PUF designs in field programmable gate array. Nonlinear Dyn. **86**(2), 975–986 (2016). https://doi.org/10.1007/s11071-016-2938-3
4. Dagadu, J.C., Li, J.: Context-based watermarking cum chaotic encryption for medical images in telemedicine applications. Multimedia Tools Appl. **77**(18), 24289–24312 (2018). https://doi.org/10.1007/s11042-018-5725-y
5. Kumar, M., Gupta, P.: A new medical image encryption algorithm based on the 1D logistic map associated with pseudo-random numbers. Multimedia Tools Appl. **80**(12), 18941–18967 (2021). https://doi.org/10.1007/s11042-020-10325-6
6. Wang, H., Tao, X., Huang, J.S.: An improved chessboard covering algorithm with generalized fractal strategy (2019)
7. Agarwal, S.: Secure image transmission using fractal and 2D-chaotic map. J. Imaging **4**(1), 17 (2018)
8. Es-Sabry, M., et al.: Securing images using high dimensional chaotic maps and DNA encoding techniques. IEEE Access **11**, 100856–100878 (2023)
9. Roy, S., Shrivastava, M., Rawat, U., Pandey, C.V., Nayak, S.K.: IESCA: an efficient image encryption scheme using 2-D cellular automata. J. Inf. Secur. Appl. **61**, 102919 (2021)
10. Elshamy, A.M., et al.: Optical image cryptosystem using double random phase encoding and Arnold's Cat map. Opt. Quant. Electron. **48**, 1–18 (2016)
11. Mohanrasu, S.S., Udhayakumar, K., Priyanka, T.M.C., Gowrisankar, A., Banerjee, S., Rakkiyappan, R.: Event-triggered impulsive controller design for synchronization of delayed chaotic neural networks and its fractal reconstruction: an application to image encryption. Appl. Math. Model. **115**, 490–512 (2023)
12. Chang, H., Wang, E., Liu, J.: Research on image encryption based on fractional seed chaos generator and fractal theory. Fractal fract. **7**(3), 221 (2023)
13. Gan, Z., et al.: Visually meaningful image encryption scheme using multi-parameter fractal theory and block synchronous sorting diffusion. Phys. Scr. **98**(8), 085216 (2023)
14. Sheela, S.J., Suresh, K.V., Tandur, D.: Image encryption based on modified Henon map using hybrid chaotic shift transform. Multimed. Tools Appl. **77**(19), 25223–25251 (2018). https://doi.org/10.1007/s11042-018-5782-2
15. Wen, H., Lin, Y., Xie, Z., Liu, T.: Chaos-based block permutation and dynamic sequence multiplexing for video encryption. Sci. Rep. **13**(1), 14721 (2023)
16. Dhingra, D., Dua, M.: A chaos-based novel approach to video encryption using dynamic S-box. Multimedia Tools Appl. **83**(1), 1693–1723 (2024)
17. Jiang, D., et al.: Real-time chaotic video encryption based on multi-threaded parallel confusion and diffusion. Inf. Sci. **666**, 120420 (2024)
18. Fouzar, Y., Lakhssassi, A., Ramakrishna, M.: A novel hybrid multikey cryptography technique for video communication. IEEE Access **11**, 15693–15700 (2023)
19. Li, C., Luo, G., Qin, K., Li, C.: An image encryption scheme based on chaotic tent map. Nonlinear Dyn. **87**, 127–133 (2017)

20. Wu, X., Zhu, B., Hu, Y., Ran, Y.: A novel color image encryption scheme using rectangular transform-enhanced chaotic tent maps. IEEE Access **5**, 6429–6436 (2017)
21. Wang, X.Y., Zhang, Y.Q., Bao, X.M.: A novel chaotic image encryption scheme using DNA sequence operations. Opt. Lasers Eng. **73**, 53–61 (2015)
22. Alanezi, A., et al.: Securing digital images through simple permutation-substitution mechanism in cloud-based smart city environment. Secur. Commun. Netw. **1**, 6615512 (2021)
23. Khairullah, M.K., Alkahtani, A.A., Bin Baharuddin, M.Z., Al-Jubari, A.M.: Designing 1D chaotic maps for fast chaotic image encryption. Electronics **10**(17), 2116 (2021)

Enhancing Cloud Security for Structured Data: An AES-GCM Based Format-Preserving Encryption Approach

S. Vidhya(✉)

Department of Computer Science, KG College of Arts and Science, Coimbatore, Tamilnadu, India
vidhya.s@kgcas.com

Abstract. Globally, 8.1 billion people exist, and among them, 5.44 billion are active Internet users that is 67.1% of global population handling vast amounts of sensitive data. This data is often structured and requires encryption. Traditional database encryption algorithms modify the format and length of the structured data, resulting in differences in size, format, and data type between the original and encrypted fields. Such modifications are particularly problematic for indexed fields, necessitating additional handling and changes in the schema of the existing table. Format-Preserving Encryption (FPE) retains the original data type and format post-encryption. This paper proposes an AES-based FPE for structured data types on the cloud. AES-GCM (Advanced Encryption Standard - Galois/Counter Mode) is a widely-used encryption method in cloud environments, combining high security with efficiency. It offers both data confidentiality and integrity, by encrypting data while also generating an authentication tag to ensure data hasn't been altered. AES-GCM's parallelizability makes it highly performant, and its non-deterministic nature, achieved through unique initialization vectors (IVs), ensures strong protection against replay attacks. It's particularly favored in cloud storage and communications for its balance of security and speed.

Keywords: Cloud security · Format Preserving Encryption · AES-GCM Encryption · Structured Data Encryption · Traditional Encryption Limitations

1 Introduction

In today's digital landscape, the secure handling of sensitive structured data is a critical requirement, especially as the number of active internet users grows exponentially. Traditional encryption methods, while providing data security, often result in significant challenges for databases. These encryption algorithms typically modify the format and length of the original data, which introduces complexities in query execution, indexing, and schema management. For instance, when encrypting an indexed field, the database schema must be altered to accommodate changes in field size and data type, which can introduce overheads in query processing and complicate database operations.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 196–205, 2025.
https://doi.org/10.1007/978-3-031-86305-9_14

Consider the example of a database containing customer account details, including credit card numbers. Before encryption, the CreditCardNumber field is stored as a VARCHAR(16). After applying traditional AES encryption, this field is transformed into a VARBINARY(256) type, causing a mismatch in data format and length. Such transformations become problematic when encrypting multiple fields, as they require significant schema adjustments, leading to inefficiencies in database management and increased system overhead.

To address these issues, Format-Preserving Encryption (FPE) has emerged as a viable solution. FPE maintains the original format and data type of the encrypted fields, eliminating the need for schema modifications and preserving the integrity of database operations. This paper proposes an efficient and secure encryption mechanism using the AES-GCM (Advanced Encryption Standard - Galois/Counter Mode) approach for structured data encryption in cloud environments. By preserving the length, format, and data type of the encrypted fields, the proposed solution ensures seamless database integration without compromising security or performance.

Creating the sql table for customers account details including the field credit card number. The following example shows how the length and datatype of the field is changed after encryption (Table 1).

```
CREATE TABLE Customers (
    CustomerID INT PRIMARY KEY,
    Name VARCHAR(100),
    CreditCardNumber VARCHAR(16)
);

INSERT INTO Customers (CustomerID, Name, CreditCard-
Number) VALUES
(1, 'Vidhya', '1234567812345678'),
(2, 'Kabilan', '2345678923456789'),
(3, 'Keerthi', '3456789034567890');
```

Table 1. Customer table before encryption

CustomerID	Name	CreditCardNumber
1	Vidhya	1234567812345678
2	Kabilan	2345678923456789
3	Keerthi	3456789034567890

After encrypting the field 'CreditCardNumber' using traditional AES encryption the filed length and datatype is changed. This is only single field encryption. If it is applying for multiple fields, the querying and indexing become too difficult and the schema is also changed drastically (Table 2).

Table 2. Customer table after AES encryption

CustomerID	Name	CreditCardNumber
1	Vidhya	0xE8F04C3B2A6F6A1F9B4F9D8C17A87F94E8F04C3B2A6F
2	Kabilan	0xA1C24F8E4B2A6D1E9B4F7A9C17A87F94A1C24F8E4B2A
3	Keerthi	0xF8E1D4B3A7F6A2F1E9B4C7D9A17A87F9F8E1D4B3A7F6

After encryption, the 'CreditCardNumber' field type is changed as VARBINARY(256) from VARCHAR(16) (Fig. 1).

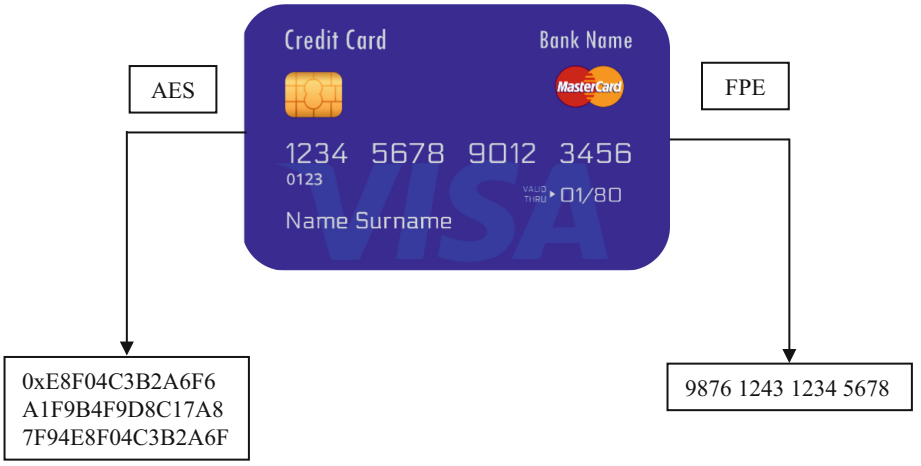


Fig. 1. AES Vs FPE Encryption

This paper's proposed solution, using AES-GCM with FPE, retains the original VARCHAR(16) format, facilitating easier integration with existing databases and reducing the impact on query processing and indexing.

2 Related Work

To secure digital data, a number of researchers have used various encryption approaches, such as a Format Preserving Encryption method by accumulating with Advance encryption standard (AES), eXclusive OR operation and a translation method for 16 digit numeric data [1].

Y. Lu, W. Xu, and H. Wang discuss the implementation of AES-GCM encryption in a parallelized manner to enhance performance in real-time cloud environments. The focus is on optimizing AES-GCM to handle large-scale data while ensuring encryption speed and security. It explores how parallelization can accelerate encryption without compromising data confidentiality and integrity, particularly relevant for cloud storage and communications. The paper also highlights AES-GCM's ability to defend against replay attacks using initialization vectors (IVs) [4].

The paper "Faster Secure AES-GCM Implementations for Cloud Environments" by S. Gueron and V. Krasnov presents optimized implementations of AES-GCM for enhanced security and performance in cloud environments. It focuses on improving the speed of AES-GCM through advanced techniques such as hardware acceleration and parallel processing. The authors emphasize the importance of balancing encryption speed with robust security, particularly for cloud computing systems, where both throughput and real-time performance are critical. The paper demonstrates how their proposed implementations can significantly reduce latency and improve encryption efficiency [5].

The paper "Efficient and Secure AES-GCM Implementation for Cloud Data Protection" by J. Lee, S. Hong, and J. Kim explores optimized AES-GCM implementations for securing data in cloud environments. The authors focus on achieving both high efficiency and strong security, leveraging techniques such as hardware-based acceleration and optimized encryption processes. The paper addresses challenges related to latency and computational overhead, proposing methods to improve encryption speed without sacrificing security. It also demonstrates the effectiveness of AES-GCM in protecting sensitive cloud data from threats like data breaches and replay attacks [7].

The reviewed studies consistently highlight the importance of optimizing AES-GCM encryption to meet the demands of modern cloud environments. Various approaches, such as parallel processing and hardware acceleration, have been proposed to enhance both encryption speed and security. These optimizations help address challenges like latency, computational overhead, and real-time performance. Furthermore, AES-GCM's ability to ensure data confidentiality, integrity, and protection against replay attacks makes it an ideal candidate for securing sensitive structured data, which aligns with the goals of this work in improving cloud data encryption efficiency and format preservation.

3 Format Preserving Encryption Modes

3.1 NIST Approved FPE Modes

The National Institute of Standards and Technology (NIST) introduced FF1 and FF3 modes for FPE. FF1, originally called FFX, and FF3 correspond to the BPS-BC component. FF1 requires an additional tweak to prevent dictionary attacks and is complex to implement. FF3, though simpler, is not suitable for general-purpose FPE due to identified vulnerabilities. The most common third generation common attacks on FF3 are A symmetric slide attack, asymmetric slide attack and slide attack using the cycle structure [2]. Both algorithms use a Feistel Network, dividing the input into halves and processing them through nonlinear functions.

3.2 Block Cipher Mode of Operation

Data security involves securing both data at rest and in transit, with encryption playing a crucial role. AES block cipher modes like CBC, ECB, OFB, CTR and CFB provide confidentiality without ensuring integrity. AES-CTR mode, suitable for FPE, transforms AES into a stream cipher, offering simplicity, parallelization, and pre-computation capabilities, while maintaining the ciphertext length equal to the plaintext. To ensure data integrity, AES-CTR can be paired with HMAC, though HMAC lacks parallelization [3].

4 Proposed Algorithm: FPE Using AES with GCM Mode

AES-GCM combines CTR mode encryption with Galois mode authentication, increasing throughput and efficiency. It processes four inputs—plaintext, secret key, IV(Initial Vector) and Authenticated Data—producing ciphertext and an authenticated tag. The encrypted fields retain the original format and data type, facilitating integration without database reconstruction [6].

- Initialization: A single, secret key is used to derive other parameters like the initialization vector (IV) and hash key.
- Encryption: Data is encrypted using the AES algorithm in counter mode.
- Authentication: The Galois field multiplication is used to generate an authentication tag that ensures the integrity of the ciphertext and associated data (Fig. 2).

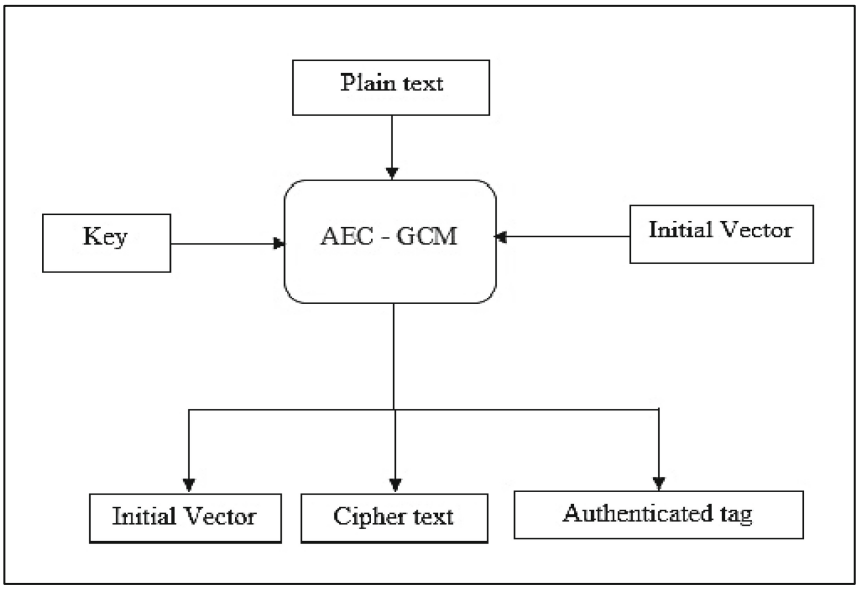


Fig. 2. AES –GCM Block diagram

- Parallelization: GCM mode supports parallel implementation, enhancing throughput.
- Efficiency: GCM uses the same code set for encryption and decryption, reducing storage requirements.
- Key Management: A single secret key is used, with the hash key and IV derived during the process, minimizing storage needs.
- Performance: GCM with 4K storage outperforms other modes for messages under 1024 bytes, and GCM with 256-byte storage is faster than CBC-HMAC-SHA1 for similar messages.
- Flexibility: Supports messages of arbitrary length without prior knowledge of the length, suitable for real-time applications requiring length preservation.

4.1 Python Script for AES – GCM Encryption and Decryption

```
import os
from cryptography.hazmat.primitives.ciphers import Ci-
pher, algorithms, modes
from cryptography.hazmat.backends import default_backend

def encrypt_credit_card(cc_number: str, key: bytes) ->
str:
    nonce = os.urandom(12)
```

```
    cipher = Cipher(algorithms.AES(key),
modes.GCM(nonce), backend=default_backend())
    encryptor = cipher.encryptor()
    ciphertext = encryptor.update(cc_number.encode()) +
encryptor.finalize()

    # Concatenate nonce, tag, and ciphertext
    encrypted_data = nonce + encryptor.tag + ciphertext

    # Convert to decimal representation
    decimal_representation = int.from_bytes(en-
crypted_data, byteorder='big')

    # Convert to a 16-digit decimal string, padded with
zeros if needed
    decimal_str = str(decimal_representation).zfill(16)[-
16:]  # Ensure 16 digits

    return decimal_str

def decrypt_credit_card(decimal_data: str, key: bytes) ->
str:
    # Convert decimal representation back to bytes
    encrypted_data = int(decimal_data).to_bytes((int(dec-
imal_data).bit_length() + 7) // 8, byteorder='big')

    # Extract nonce, tag, and ciphertext
    nonce = encrypted_data[:12]
    tag = encrypted_data[12:28]
    ciphertext = encrypted_data[28:]

    cipher = Cipher(algorithms.AES(key), modes.GCM(nonce,
tag), backend=default_backend())
    decryptor = cipher.decryptor()

    try:
        return decryptor.update(ciphertext) + decryp-
tor.finalize().decode()
    except Exception as e:
        print(f"Decryption error: {e}")
        return None

# Example usage
if __name__ == "__main__":
```

```
    key = os.urandom(32)  # AES key should be 16, 24, or
32 bytes long
    cc_number = "1234567812345678"  # Example credit card
number (16 digits)

    encrypted_decimal = encrypt_credit_card(cc_number,
key)
    print(f"Encrypted (Decimal, 16 digits): {en-
crypted_decimal}")

    decrypted_cc_number = decrypt_credit_card(en-
crypted_decimal, key)
    print(f"Decrypted: {decrypted_cc_number}")
```

PS C:\Users\Admin\Downloads> c:; cd 'c:\Users\Admin\Downloads'; & 'c:\Users\Admin\AppData\Local\Programs\Python\Python312\python.exe' 'c:\Users\Admin\.vscode\extensions\ms-python.debugpy-2024.10.0-win32-x64\bundled\libs\debugpy\adapter/../..\debugpy\launcher' '54560' '--' 'C:\Users\Admin\Downloads\GCMAESPY.py'

Encrypted (Decimal, 16 digits): 0291038984510601

Fig. 3. Output of the above script

The plaintext and ciphertext are same in length and format. To check the integrity the cipertext is concatenated with nonce and tag (Fig. 3).

5 Results and Discussion

AES – GCM mode is more suitable for FPE due to the following reasons.

- Authenticated Encryption: AES-GCM provides both encryption and authentication (integrity) in a single operation. It ensures that data is both confidential and has not been tampered with. This makes it particularly suitable for securing data in transit.
- Performance: AES-GCM is generally faster and more efficient than FF1 and FF3 due to its use of parallelizable operations. The Galois field multiplication and counter mode operations can be efficiently implemented in hardware and software.
- Standardization: AES-GCM is widely adopted and standardized (e.g., NIST SP 800-38D). Its usage is supported in many cryptographic libraries and standards, making it a reliable choice for many applications.
- Flexibility: AES-GCM can handle variable-length data, making it versatile for various use cases beyond format-preserving encryption.

- Security: AES-GCM is based on AES, which is a well-established and highly secure encryption algorithm. It provides strong security guarantees when used with a unique nonce for each encryption.

AES-GCM does not preserve the format of the plaintext. Encrypted data may have a different length and structure compared to the original data. But this issue is simply overcome by finally converting the ciphertext to require representation. The only thing to be ensured nonces are unique and not reused is critical. Reusing a nonce with the same key can compromise security [8] (Table 3).

Table 3. Performance analysis of AES –GCM for FPE

Features	Traditional encryption	FF1	FF3	AES-GCM
Length Preservation	No	Yes	Yes	Yes
Format Preservation	No	Yes	Yes	Yes
Index usability	No	Yes	Yes	Yes
Database restructuring	No	Yes	Yes	Yes
Algorithm complexity	Low	Moderate	Moderate	High
Parallelization	No	No	No	Yes
Data integrity	No	No	No	Yes
Storage requirements	Low	Moderate	Moderate	High
Scalability	Low	Moderate	High	High
Security Level	Moderate	High	High	Very High
Nonce Requirements	Not applicable	Not applicable	Not applicable	Unique required
Performance in Real-Time	Limited	Moderate	Moderate	High

This table compares traditional encryption, FPE, and the proposed AES with GCM mode based on key performance metrics, highlighting the benefits of using AES-GCM for structured data encryption.

6 Conclusion

Format-Preserving Encryption (FPE) addresses these challenges by retaining the original data format and type after encryption, thereby reducing the need for extensive schema adjustments and preserving referential integrity. The proposed AES-based FPE algorithm offers an effective solution for encrypting structured data in cloud environments, combining the benefits of encryption with the advantages of format preservation. Additionally, its ability to be implemented in parallel enhances encryption speed, making it a practical choice for managing large volumes of sensitive data efficiently.

Overall, the proposed AES-based FPE algorithm provides a balance between data security and operational efficiency, addressing the shortcomings of traditional encryption methods while facilitating seamless integration into existing systems.

References

1. Boneh, D., Dingledine, R., Shoup, V.: A new format-preserving encryption scheme. In: IEEE Symposium on Security and Privacy (S&P) 2021, pp. 317–334 (2021)
2. Amon, O., Dunkelman, O., Keller, N.: Three third generation attacks on the format preserving encryption scheme FF3. In: Canteaut, A., Standaert, FX. (eds.) EUROCRYPT 2021, LNCS, vol. 12696, pp. 127–154, Springer, Cham (2021). https://doi.org/10.1007/978-3-030-77886-6_5
3. Cramer, R., Ducas, L., Wesolowski, B.: A tight security analysis of FPE schemes in the quantum random oracle model. In: Advances in Cryptology – CRYPTO 2022. LNCS, vol. 13182, Springer, Cham (2022)
4. Lu, Y., Xu, W., Wang, H.: Parallelized AES-GCM for real-time data encryption in cloud computing. IEEE Trans. Cloud Comput. **9**(2), 389–398 (2021)
5. Gueron, S., Krasnov, V.: Faster secure AES-GCM implementations for cloud environments. In: Proceedings of the 2021 IEEE Symposium on High-Performance Computer Architecture (HPCA 2021), pp. 478–489 (2021)
6. McCurley, K.S., Ross, J.B.: Recent advances in format-preserving encryption. In: Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS 2022), pp. 845–860 (2022)
7. Lee, J., Hong, S., Kim, J.: Efficient and secure AES-GCM implementation for cloud data protection. Int. J. Inf. Secur. **20**(3), 451–467 (2021)
8. Boldyreva, A., Chenette, N., O'Neill, A.: A security analysis of format-preserving encryption in the frequency domain. In: Cryptology ePrint Archive, Report 2021/1234 (2021)

Hacktivism or Cyber Warfare? Decoding the Motivations Behind Cyber Attacks Targeting Israel

Niraj Kumar Singh(✉) and Sunil Bajeje

Marwadi University, Rajkot, India
nirajsingh98456@outlook.com

Abstract. The realm of cybersecurity is increasingly fraught with sophisticated threats, with cyber attacks targeting nations often shrouded in ambiguity regarding their underlying motivations. This study delves into the intricate landscape of cyber assaults directed at Israel, aiming to disentangle the complex web of motives driving such attacks. Through an extensive analysis of diverse cyber incidents and the examination of associated ideological, political, and strategic underpinnings, this research endeavors to delineate between hacktivist endeavors and manifestations of cyber warfare. By scrutinizing the historical context, tactics, and messaging deployed by threat actors, this study seeks to offer clarity on the underlying intentions behind cyber operations directed at Israel. The results of the study indicate that the conflict between Palestine and Israel has led to a cyber warfare landscape involving other nations and various hacktivists from different countries supporting both sides. Israel's cybersecurity weaknesses have allowed various hackers and hacktivists to exploit opportunities with motives such as political, economic, business, ideological interests, self beliefs. In response, Israel established the National Cyber Directorate, responsible for cybersecurity policy, security, and capacity development for the government, public, and private sectors, to prevent and mitigate threats to critical infrastructure and manage cyber incidents. Additionally, Israel engages in cybersecurity cooperation with various countries to combat cyber terrorism.

Keywords: Cybersecurity threats · Cyber attacks on Israel · Hacktivist endeavors · Cyber warfare · Motivations behind cyber operations · National Cyber Directorate

1 Introduction

In the realm of cybersecurity, Israel is internationally recognized as a formidable stronghold renowned for its expertise and leadership in safeguarding digital infrastructure. Undoubtedly, Israel is widely acknowledged as a powerhouse in the realm of cybersecurity, known for its competence and leadership in safeguarding digital infrastructure. Israel is recognized as a nation with unmatched worldwide cybersecurity expertise. Israel has effectively showcased its capacity to safeguard digital assets and infrastructure from

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 206–232, 2025.
https://doi.org/10.1007/978-3-031-86305-9_15

global cyber threats. There is no longer any doubt about his robust security capabilities from Siberia. Nevertheless, Israel, known for its robust standing in the field of cybersecurity, confronts a multitude of obstacles and susceptibilities due to warfare and hacktivism that specifically aim at its cyber infrastructure and national interests.

Israel is confronted with a multitude of intricate and varied cyber threats, which poses a significant challenge. Israel frequently faces cyber attacks from a range of players, including rival nations, violent factions, and individual or hacktivist groups. The attacks encompass DDoS (Distributed Denial of Service) attacks, data breaches, intrusion into government or business networks, as well as online propaganda and disinformation efforts. Furthermore, Israel is facing significant challenges in safeguarding its extensive and multifaceted cyber infrastructure.

In the era of digitalization, where technology has advanced and the Internet is widely used, safeguarding vital systems such as military infrastructure, governmental services, and private sector networks has become progressively intricate for all nations. Despite Israel's impressive technological capabilities, it remains vulnerable due to the constantly evolving and expanding landscape of cyber threats worldwide. The frequency of cyber attacks is increasing due to the use of novel methods that exploit recently identified weaknesses in software or systems, as well as the manipulation of human susceptibility through social engineering strategies.

The enduring struggle between Israel and Palestine transcends the physical battleground and extends into the digital domain, influencing the worldwide perception of leadership and the capacities of cybersecurity. Evaluating the resilience of national security and vulnerabilities in the current era heavily relies on analyzing the cyberwar scene. Conventional armed conflict and cyber warfare give rise to a varied and ever-changing setting where participants from both sides employ techniques that span from activism to intricate hacking maneuvers. During the armed confrontation between Israel and the Palestinians on October 7, 2023, a notable cyber attack was directed at the "Noga Independent System Operator," which had the responsibility of overseeing the management of the Israeli electricity grid. The incident underscored the interconnectedness of physical and digital security risks, emphasizing the possible consequences for national stability and the integrity of infrastructure.

Israel and Palestine are utilizing internet channels to disseminate misinformation and carry out cyber attacks amidst increasing internal tensions. The increasing occurrence of Distributed Denial of Service (DDoS) assaults, which are designed to disrupt important websites and emergency infrastructure, highlights the susceptibility of digital assets during times of war. Moreover, the utilization of internet platforms to control public discourse, together with the dissemination of pro-Israeli or pro-Palestinian content, adds further complexity to the story of the Israeli-Palestinian conflict. In addition, reputable sources, such as Microsoft, have emphasized the participation of cyber organizations like "Storm-1133" located in Gaza, which successfully penetrated Israeli telecommunications businesses using advanced cyber espionage techniques. Furthermore, the utilization of social engineering techniques, such as fabricating counterfeit LinkedIn profiles portraying Israeli HR managers, with the intention of disseminating malware among employees of Israeli organizations, highlights the tactics devised by adversaries.

The repercussions of cyber assaults extend beyond their direct targets, affecting numerous government websites, media outlets, financial institutions, telecommunication firms, and energy companies through DDoS attacks and destructive techniques. This underscores the widespread impact of cyberattacks across public and private sectors. The involvement of hacktivist groups from various international locales, supporting Palestine, adds a transnational dimension to the conflict, underscoring the global ramifications of regional disputes and interconnected cyber threats. In light of these developments, this research seeks to elucidate the multifaceted motivations behind cyberattacks targeting Israel amid the Israel-Palestine conflict. By analyzing the tactics, actors, and consequences of these attacks, this study aims to discern whether these actions align more with hacktivism or cyber warfare, offering insights into broader implications for international cybersecurity and geopolitical stability. This research uses three concepts to analyze data according to the needs of the researcher.

1.1 Attribution Theory

Attribution theory in cybersecurity focuses on identifying and attributing cyber attacks to specific actors or entities. This theory can help researcher analyze the motives behind cyber attacks targeting Israel by examining the characteristics and behavior of the attackers. By studying patterns of cyber activity, tactics, and known affiliations of threat actors. Utilizing this theory enables researchers to gain insights into whether the attacks are more aligned with hacktivist motivations or state-sponsored cyber warfare strategies.

1.2 Cyber Warfare vs. Hacktivism

The research aims to distinguish between cyber warfare and hacktivism. Understanding the characteristics, tactics, and objectives of each can help in categorizing and analyzing the cyber attacks directed at Israel.

1.3 Cyber Deterritorialization

Cyber deterritorialization is a concept that expands the scope of future warfare security by distinguishing between traditional and modern warfare frameworks. With the advent of digitalization, warfare is no longer limited to military conflicts, nuclear weapons, or chemical weapons; it also includes the potential use of modern warfare such as cyber warfare, which can be wielded by both state and non-state actors. The concept of cyber deterritorialization views traditional threats to states as being very limited to warfare conducted solely by states. However, the widespread use of cyber capabilities among the global community has increased the use of cyber weapons that can jeopardize international stability and security. Therefore, to maintain the stability and security of the country, all stakeholders in various fields require prevention against cyber attacks, cyber crime, and cyber warfare.

2 Related Work

As per paper [10] the research explores the relationship between hacktivism and website defacement, as well as its impact on companies and government organizations. Using qualitative analysis of hacktivists' motives and modus operandi, the study finds that they have diverse ideological and psychological motivations. They often exploit known vulnerabilities and techniques, as well as publicly available tools. Targets are selected based on their ease of compromise and the potential attention the defacement may receive. The research involves literature review, interviews with hacktivists and cybersecurity experts, and forensic data analysis. Although many questions remain unanswered, this study provides valuable insights into the phenomena of hacktivism and website defacement.

As per paper [11] the journal "Israel-Palestine Cyber Conflict" by Mohammed Abu Saada provides a comprehensive overview of the cyber conflict between Israel and Palestine. The researcher highlights the importance of cyberspace in this conflict, used for recruitment, propaganda and psychological warfare by various groups and organizations, especially global jihadist groups since 2012. The journal describes the development of communication and information technology in Palestine, including the establishment of institutions such as the Ministry of Communications and Informatics and the Palestinian Information and Communication Technology Incubator. The internet in Palestine has become a means to reduce unemployment and connect with the world, although their cyber infrastructure still depends on Israel. Israel is described as a country with significant advances in technology and cyberspace since the 1990s, with cyber innovations generating diplomatic and political advantages. The journal also discusses tools of cyber conflict between Israel and Palestine, such as hacking and social media, with Israel having an advantage However, the Palestinian cyber team is also capable of carrying out effective attacks so that cyber attacks between the two parties will likely continue to increase.

As per paper [12] the paper "From Kinetic to Cyber Attacks and Back: The Israeli Approach to Deterrence in Cyberspace and the Multi-Dimensional Threat of Hamas" by Beatrice Gori, analyzes Israel's use of cyber tools to counter threats from Hamas and other non-state actors in the context of the Israeli-Palestinian conflict. The conflict has expanded into cyberspace, leading to a shift in Israel's security strategy towards an offensive-deterrent posture. This has resulted in escalating insecurity, as seen in the May 2021 clashes. The paper highlights the challenges posed by the spread of power struggles into cyberspace and emphasizes the need for a redefinition of the concept of war to account for cyberspace's unique characteristics. Overall, the paper underscores the importance of understanding the evolution of the conflict and the role of cyber tools in developing effective security strategies in the region.

As per paper [13] article "Deterritorializing Cyber Security and Warfare in Palestine: A Literature Review" by Fabio Cristiano's, challenges the assumption that national cyberspace reflects a country's territorial sovereignty, particularly in the context of the Occupied Palestinian Territory (OPT). The article examines the intersecting cyber security strategies of the Israeli government, the Palestinian Authority, and Hamas, highlighting how these strategies produce a cyberspace characterized by territorial annexation, occupation, and blockade. By employing the conceptual prism of (de-)–(re-)

territorialization, the article reflects on how these strategies, as well as those of Palestinian hackers, articulate territoriality beyond the normativity of national cyberspace. Cristiano's work provides a critical perspective on the relationship between national cyberspace and territorial sovereignty in the Israeli-Palestinian conflict.

3 Cyber Attack on Israel

David Shamah in his work asserts that Israel, known as the start-up Nation is currently home to over 3,500 new technology companies, with one company for every 1,400 Israeli citizens. Indeed, Israel boasts the highest number of start-ups per capita in the world. Israel's dominance in cybersecurity is particularly significant. In 2017, the National Cyber Array of Israel reported over 6,500 cyberattacks on Israel daily, representing a 60% increase from 2016. Globally, Israel is recognized as the second most targeted country for cyberattacks, trailing only the United States. This is facilitated by Israel's other technological prowess, namely its "national emphasis on the use of Artificial Intelligence".

The conflict between Israel and Palestine extends beyond physical confrontations to include cyber warfare. This is evident in the increasing emphasis on cyber attacks targeting various security sectors, including government institutions, with a notable 52% rise in such attacks. These cyber operations employ tactics like Distributed Denial of Service (DDoS) attacks, malware distribution, and other vulnerability exploits. These aim to disseminate misinformation, particularly regarding potential rocket attacks, in order to instill fear among the Israeli population. Israel, recognized for its robust cyber security capabilities, responds to these cyber attacks, prompting questions about the possible support for these attacks solely by HAMAS or if there is involvement from other pro-Palestinian states with a more profound understanding of the cyber domain, such as Iran.

On October 2 to 10 2023, it was recorded that Israel was one of the countries that experienced the most DDoS attacks with websites attacked 143 times. Many Israeli websites are targeted by the Hamas group, pro-Palestinian activists and also in several cases handled by hacker activists from Russia (Fig. 1).

On October 2 and 10, researchers from Radware discovered that Israel was targeted by 143 DDoS attacks, claimed by 43 hacktivists on Telegram. On October 7, during an attack by Hamas on Israel, there were also 30 DDoS attacks, claimed by 40 hacktivists. After the Hamas attack on Israel on October 9, there was a DDoS attack on the Re Alert app, which notifies civilians of rocket attacks. The initial attack peaked at 100,000 requests per second (rps) and lasted ten minutes. Forty-five minutes later, a second, much larger attack peaked at 1 million rps. In addition to the DDoS attacks, the hacker group AnonGhost exploited vulnerabilities in the Red Alert app to intercept requests, expose servers and APIs, and send false alerts to users with messages like "a nuclear bomb is coming."

On October 14, 2023, Cloudforce One Threat Operations identified an Android app that mimicked the Red Alert app. This app accessed sensitive information such as phone contacts, SMS, call logs, and SIM card information. The second major target was the computer software industry, with nearly 34% of all DDoS attacks targeting

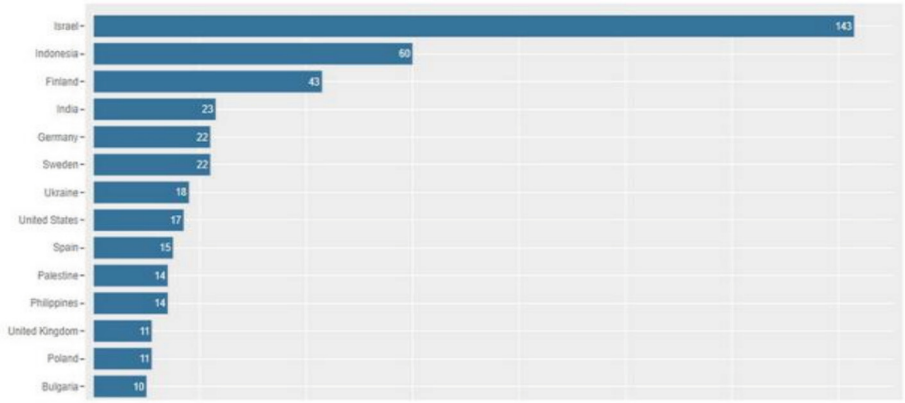


Fig. 1. World's Highest Ransomware Attack

software companies. The third target was companies in the Banking, Financial Services, and Insurance (BFSI) sector, followed by government administration websites in fourth place. Additionally, an attack on The Jerusalem Post's website on October 9, 2023, caused the site to go down, with 58 hacking groups claiming responsibility.

According to FalcFeeds.io, there are over 100 groups engaged in attacks against Israel. Among them, approximately 70 are supportive of the Palestinian cause, while around 20 favor Israel. The remaining groups maintain a neutral stance, targeting both sides. Those aligned with the Palestinians are reportedly linked to countries like Iran and Russia, known for their involvement in cyber attacks against Israel. Additionally, some of the groups suspected of being hackers are located in Malaysia and Indonesia.

The hacker's main target was 36%, targeting Israeli government websites, news and media 10%, then travel 9%. While financial websites accounted for 5.6%, education 4.2% and healthcare 3.5%. DDoS attack actors claiming attacks on Israeli websites are Garnesia_Team (14%), Maroccoan black cyber Army (6%), Ganosec Team (8%), Mysterious Team Bangladesh (6%), Team herox (6%), Electonic_Tigers_Unit (6%), Mr.Dempsey (4%, Anonymous Sudan (4%), AC3HCYBER (4%), Sylhet Gang-SG (4%), and (36%) are other cyber attacks. Hacker groups 36% mostly claimed by actors: Indonesian threat actor Garnesia_Team, Mysterious team Bangladesh and Indian Group Team Herox. The attack was politically, ideologically and religiously motivated with support for brothers in Palestine. On the other hand, Killnet as a pro-Russian hacker group also intervened in hacking Israeli government sites and banks, such as Shabak.gov.il.

The impact of these cyber attacks is evident in their targets, which include newspapers, media outlets, banking software companies, financial and insurance services, as well as government administrative sites. For instance, a successful DDoS attack on the Jerusalem Post website rendered it inoperable for a period of two days. Notably, the attribution of cyber attacks can be complex, with various hacker groups claiming responsibility. For example, the hacking of the Jerusalem Post website was attributed to groups like "Team_ Insane_ Pakistan" and "Anonymous Sudan," which are known for their pro-Russian affiliations. Additionally, hackers leaked personal information of students

from "Ono Academic College," leading to the temporary shutdown of the university's website.

In addition, the cyber attacks targeted the Israeli rocket warning application "Red Alert", which is used to warn Israeli citizens about the dangers of a forthcoming rocket attack. The Israeli-Palestinian conflict is spreading into the cyber-welfare domain, marked by increasing cyber attacks targeting various sectors in Israel, ranging from DDoS attacks to the distribution of malware aimed at spreading fear and misinformation among the Israeli population. The assumptions related to Israel's strength and capabilities in cybersecurity are considered very strong and capable of responding effectively to various cyber threats in various sectors. However, the incidents have revealed a cyber security vulnerability that has resulted in greater involvement of actors from both non-State actors and pro-Palestinian states in organizing cyber-attack operations. It shows that cybersecurity is very complex in a digital battlefield characterized by numerous factions engaged in hostile activities with various motives and targets of attack.

In mid-October, as the attacks between the two countries intensified, both through military strikes and more damaging cyberattacks such as the spread of wiper malware and targeting Industrial Control Systems (ICS), there were also mass influence campaigns in support of Palestine aimed at creating confusion and undermining support for ground operations. Then, hackers supported by other countries such as Iran, Pakistan, and Bangladesh employed an operational approach with a surge of wiper malware, ransomware, and the deployment of cellular spyware against the Israeli government in various critical infrastructures such as the financial sector, technology, and the defence sector. The tactic was used to sabotage Critical National Infrastructure (CNI) and to gather intelligence to strengthen the state supported by hackers. Iran, as a pro-Palestine country, has launched a phishing campaign against Israeli national security think tanks, diplomats, non-governmental organizations (NGOs), and Israeli decision-makers. In addition, hackers supported by Iran are also sending mass SMS and emails and using AI to create campaign claims supporting Palestine against Israel. Furthermore, information manipulation regarding Israeli citizens' support for Palestinian independence is widely disseminated by hacktivists, creating concern among the Israeli public, especially as this information targets military personnel and Israeli officials, with photos and support templates spread through social media.

The Israeli-Palestinian conflict transcends the national boundaries between the two states alone but also involves the global power of both state and non-state actors from various countries. Additionally, hacktivism involvement in cyber attacks targets various critical infrastructures, military installations, governments, and private entities in Israel da Palestine Pro-Palestinian hacker activists have recently targeted countries affiliated with Israel such as the United States, Britain, Germany, India, France, Italy, Canada, among others. Similarly, pro-Israel hackers have targeted their attacks on pro-Palestinian countries such as Iran, Iraq, Qatar, Saudi Arabia, and Lebanon. That creates a global cyber welfare landscape in the world.

4 Analysis of the Motives Behind Cyber Attacks

4.1 Ghost of Palestine

Ghosts of Palestine, which is a group focused on supporting Palestine, claimed to have attacked the site govforms.gov[.]il, which is associated with the Israeli Ministry of Education, on October 19, 2023. The group also claimed to have hacked into the Israeli government website on October 20, 2023. The motivation behind the attack is a call for humanity in Gaza. The Ghost of Palestine is also campaigning for all pro-Palestinian hacktivists to carry out attacks on the public and private infrastructure of Israel and the United States.

4.2 Garnesia_Team

Team Garnesia is widely recognized as a prominent and aggressive cyber entity on the global stage, with its cyber capabilities and experience posing a potential threat to international stability as it aims to inflict widespread damage on national security. The group's attack on Israel was primarily motivated by economic interests, specifically aimed at weakening Israel's industrial technology output. Israel, with a robust economy boasting a GDP of $272 billion and ranking as the 40th largest economy globally, has established itself as a leader in technological advancement.

Sponsored by MarsCorp, Garnesia Team seeks a technological breakthrough in missile systems, targeting Israel's extensive technology industry with the goal of acquisition. Additionally, the attackers aim to achieve political leverage by influencing public opinion, thereby creating opportunities for governmental change. Cyber attacks are facilitated by a domestic environment where challenging the government is met with fear of counterattacks. Consequently, cyberattacks yield political and psychological benefits for the attackers, gradually eroding democratic values and human rights as societal fear and insecurity mount.

In today's modern era, society's reliance on information media as a tool to shape public opinion is a crucial economic asset. Therefore, one of the attackers' motives is to influence the attitudes and opinions of the general public, with the aspiration of becoming a political opposition force. As public opinion and domestic support shift away from the government, it will face escalating challenges in managing growing pressures. Over time, mounting dissatisfaction can evolve into a political movement capable of effecting change in government.

4.3 Maroccoan Black Cyber Army

The Moroccan Black Cyber Army is a prominent group within the dark web, comprising not only hackers from Morocco but also individuals with backgrounds from various Middle Eastern and African countries, including Algeria, Mauritania, Libya, and Nigeria. On January 13, 2020, the group launched a large-scale cyber attack on Israel, driven by its broader anti-imperialist agenda. The motivation behind these cyber attacks involves political, religious, and ideological factors.

Politically, the Moroccan Black Cyber Army conducts cyber attacks against Israel to express solidarity with Palestine and to oppose what they perceive as unjust Israeli policies. They aim to use cyber attacks as a tool to harm Israel both economically and politically.

Religiously, the group's motivation stems from the significance of Al-Aqsa as a holy site for Muslims. They believe that by targeting Israel through cyber attacks, they can contribute to the defense and protection of the holy site from perceived Israeli threats.

Ideologically, the Moroccan Black Cyber Army is motivated by their belief in the importance of achieving independence and justice for the Palestinian people. They utilize their expertise and technological capabilities to combat what they perceive as Israeli oppression of Palestinians.

The cyber attacks employed by the group typically involve Distributed Denial of Service (DDoS) attacks, which overload the target's servers with excessive web traffic, causing them to crash and go offline. These attacks affected over 40 government and private Israeli websites, including critical institutions such as the parliament, central bank, finance ministry, immigration authorities, cell phone providers, leading national newspapers, and smaller local companies. Besides the direct impact of website shutdowns, these attacks also have indirect digital consequences for users and businesses.

4.4 Ganosec Team

The Ganosec team emerged in the mid-2010s, amidst the backdrop of the Arab Spring, and has since become known for its global cyber attacks. According to the Chief Operating Officer of security firms FireEye and Mandiant, the behavior of Ganosec does not resemble that of an espionage operation but rather indicates a motive for hacking for financial gain, suggesting that their activities are not driven by a desire to support the Palestinian cause but rather to profit on the black market.

In April 2019, the United States Cybersecurity and Infrastructure Security Agency issued a warning regarding the hacking of the Israeli Ministry of Energy, linking it to Ganosec. Furthermore, the Ganosec team's use of distributed denial of service (DDoS) attacks aligns with the profit-driven nature of their operations, as DDoS attacks aim to render online services unavailable by overwhelming them with traffic from multiple sources. Victims of such attacks incur costs for filtering out this traffic, providing an incentive for attackers to maximize resource consumption and thereby increase profits.

The primary objective of Ganosec's attack was to instill feelings of anger, fear, and trauma within Israeli society. Additionally, the team believed that by executing a large number of attacks, they could provoke neighboring countries into launching high-profile attacks against Israel, potentially leading to widespread conflict worldwide. Ganosec also aimed to undermine Israel's credibility in protecting its cyber systems, thereby deterring international business with the country due to concerns over data security. Moreover, Ganosec believed that the widespread global animosity towards Israel would amplify criticisms of the country's cybersecurity weaknesses, further exposing vulnerabilities in its defenses.

4.5 Mysterious Team Bangladesh

The Mysterious Team Bangladesh is a hacktivist group that engages in cyber attacks against India and Israel, driven by political and religious motivations. Their targets primarily include logistics organizations, governments, and the financial sector in these countries. In June 2022, the group launched a series of attacks, including 750 Distributed Denial of Service (DDoS) attacks that flooded websites with junk traffic and 70 attacks that altered the appearance of websites with unauthorized content. Approximately 30% of their attacks were aimed at India, while 18% targeted Israel. The Mysterious Team Bangladesh has been actively involved in cyber attacks such as DDoS attacks and website defacement since June 2022. Their motivations appear to be rooted in religious and political ideologies, as evidenced by their online presence and statements. The group also targets countries like Sweden, often in response to specific events or news triggers. The emergence of groups like the Mysterious Team Bangladesh could be linked to escalating geopolitical tensions, with hackers seeking recognition and potentially monetizing their activities through advertising.

4.6 Mr. Dempsey

Mr. Dempsey, analyzed by Resecurity based on their claims of attacks and motives, is a cyber threat actor who identifies themselves as the Systemadmindbd Spy Agency from Bangladesh, with a motive to advocate for Palestinian independence. Mr. Dempsey engages in hacktivist activities by amplifying the #FreePalestine PSYOP efforts on various social media platforms such as Facebook. Additionally, Mr. Dempsey expands the recruitment campaign of hacktivist groups from Bangladesh, Malaysia, and Pakistan, which are countries with the largest Muslim populations.

This hacking group frequently targets other countries, with India being one of their victims, attacking government and financial institutions through website defacements and DDoS attacks. On October 4, 2023, Mr. Dempsey announced an initiative against the Saudi Arabian government and a cyber attack campaign related to #FreeSaudiArabia. Subsequently, on October 6, a DDoS attack by Mr. Dempsey targeted Saudi Arabia's public transportation system. These campaigns and hacks are driven by the belief that the Saudi government is unjust due to its political ties with Israel. In essence, Mr. Dempsey's attacks aim to reinforce the geopolitical narrative regarding the support between Palestine and Israel.

4.7 Killnet

Hackers who claim to be patriotic Russian volunteer hackers. Killnet openly announced that they are targeting attacks against the Israeli government system with DDoS attacks that flood websites and force them offline. Killnet launched an attack against Israel with the motive of blaming Israel for the conflict between Russia and Ukraine due to Israel's support for Ukraine's NATO membership. The involvement of Killnet as patriotic Russian volunteer hackers utilized DDoS cyber attacks on Israeli government systems to convey the message that Russia supports Hamas and opposes Israel in the Israel-Palestine conflict.

4.8 Anonymous Sudan

The Anonymous Sudanese hacktivist group, associated with Russia, demonstrates solidarity with Palestinian Hamas. The group's primary focus is disrupting Israel's Industrial Control Systems (ICS) and Supervisory Control and Data Acquisition (SCADA) systems, presenting a significant cybersecurity risk. Their campaign was announced through their Telegram channel, detailing specific targets within Israel's Global Navigation Satellite System (GNSS), Building Automation and Control Network (BACnet), and Modbus Industrial Control Systems.

Originating from a Russian-language Telegram channel in January 2023, the group identifies as Anonymous Sudan, despite their lack of affiliation with Sudan, indicating a stronger connection to Russia. The Anonymous attackers are closely associated with other Russian hacktivist groups like Killnet, which are famous for their DDoS attacks. These alliance groups highlight networked cyber groups that are interconnected by the same ideologies. Sudan Anonymous frequently uses DDoS attacks to attack targets marked by a flood of UDP and SYN waves directed to the victim's servers, causing service disruption or system slowdown. In addition, these Sudan Anonyms also use free and open public cloud servers and proxy infrastructure to confuse the source of their attacks (Fig. 2).

Country	Date	Target	Motive	Impact
Sweden	08/02/23	Airline	Respond to the burning of the Al-Quran at the Turkish embassy	DDos attacks against 21 airlines
	10/01/23	Hospital	Revenge of burning the Al-Quran	The Swedish health sector has been dropped
	11/02/23	University		26 Universites
	12/02/23	Bank		16 Bank (Bank crisis)
Denmark	02/04/23	Airline	Revenge of burning the Al-Quran	DDos attacks against 11 airlines.
Australia	24/03/23	Airports	Retaliation for the provocative actions of Muslims with the word "Allah" on the naked model design.	Infrastructure is not functioning temporarily
	02/04/23	Hospital		DDoS attacks cause patients to be unable to be treated because the system is not functional
France	15/03/23	Flight company	Revenge for cartoons containing the prophet Muhammad naked	DDoS attacks on 5 airlines caused flights to be disrupted.
	16/03/23	Corporate & healthcare sector		Damage to the infrastructure of 9 French companies in several hospitals caused the hospitals to be unable to accept patients.
	17/03/23	Universities		DDoS attacks cause university infrastructure to collapse
Israel	02/04/23	Hospital	Respond to the Israeli authorities' violent actions at the Al-Aqsa Mosque	
	04/04/23	Schools	Respond to Israel's treatment of Palestine	Israel education sector has been dropped because of DDoS attack
	04/04/23	Airlines, Banks, Israeli National Intelligence Organization, Universities, Newspapers.	Condemnation of Israel's treatment of Palestinians	DDoS attacks on this infrastructure and Anonymous Sudan's call for hackers with the same aim of attacking Israel for Israel's actions over the Palestinian-Israeli conflict
India	07/04/23	Airport	Revenge of what Indian doing to Muslims	6 Airports collapse because of DDoS attacks
	10/04/23	Hospitals		10 Indian hospitals are damaged and cannot be used for a certain period of time
	10/04/23	Indian E-commerce Companies		3 Indian E-commerce Companies hacked and some data be leaked
United States	21/04/23	Twitter	Attack on American support for Israel	The attack of DDos maded the page down and become temporarily unusable
	25/04/23	Hospitals	Carrying out attacks on America like Anonymous Sudan did on Israel	Several attacks on the United States' health infrastructure
Germany	28/04/23	Infrastructure in several regions in Germany	Attacks on racism against Muslims	DDoS attacks on Infrastructure in Germany including German Federal Intelligence, Frankfurt - Hahn Airport & Stuttgart Airport.
	29/04/23	Federal Police Germany & Deutsche Bank AG	The result of kidnapping children from Muslim families	Attack on the German Federal Police & Deutsche Bank AG in response to the kidnapping of children from Muslim families

Fig. 2. Anonymous Sudan beyond Hacktivist Attacks

The table above explains that the motivations of the Anonymous Sudan vary widely, ranging from attacks and cyberattacks resulting from religious disobedience to Russian and Ukrainian involvement in warfare followed by strategic cyber warfare support targeting Israel and its allies in the Israeli-Palestinian conflict. Moreover, the involvement

of Anonymous Sudan, which is regarded as hacktivism for the Palestinians in claiming responsibility for these attacks, emphasizes protest against the political, military, and economic support felt by the countries that belong to Israel. This is because the attacks are targeted not only to Israel but also to countries affiliated with Israel, which can be linked to cyber attacks against countries such as the United States, Germany, and Australia as a form of hacktivist demonstration in the cyber realm aimed at claiming power and influence in the digital world. These attacks also serve as a response to grievances concerning the foreign policies of these nations, further highlighting the complex interplay of ideological, strategic, and retaliatory motives driving Anonymous Sudan's cyber activities. In addition, Anonymous Sudan also took a stance in hacking with political motives by attacking Kenyan organizations on October 15, 2023, in retaliation for Kenya's support of Israel.

4.9 Cyber Av3ngers

This Hacktivist pro-Hamas group is targeting Israel's Independent System Operator (Noga), a power network, with threats to shut down its website. Cyber Av3ngers also targets the Israel Electric Corporation as the largest electricity supplier in Israel and the Palestinian territories. The hacktivist organization Cyber Av3ngers, aligned with Iranian interests, has engaged in cyber attacks against numerous water utility facilities in Israel. The initial motivation behind these attacks was linked to the broader Israel-Hamas conflict. The hackers perceived their actions as supportive of the Hamas group, backed by the Iranian government in the Israeli-Palestinian conflict. However, over time, the motivations of these cyber criminals have evolved and diversified, encompassing financial gain, personal ego, entertainment, political agendas, and social affiliations (Fig. 3).

4.10 Dark Storm Team

The Dark Storm Team is a hacker group that claims to be pro-Palestine, but their attacks are not limited to Israel; they also target countries outside of Israel. On October 8 and 9, 2023, this group claimed to have attacked Israeli hospitals, and they subsequently targeted the Israeli government and sensitive Israeli industries. On October 15, 2023, Dark Storm Team claimed to have attacked John F. Kennedy International Airport in New York, citing that the airport was filled with Zionist entities. The message conveyed in the announcement of the New York airport attack was that Dark Storm Team would target countries that are pro-Israel. Actually, the targets and objectives of this group are not only pro-Palestine, but when looking at their claims regarding attacks by other countries such as NATO, where on August 23, 2023, an attack targeted Denmark, and on September 10, it targeted the Netherlands, this hacktivist group is also taking a pro-Russia stance. In addition, the motives of this group are not only political support for Palestine in the Israel-Palestine conflict and support for Russia in the Russia-Ukraine conflict as a push for geopolitical orientation for both Russia and Palestine. However, behind the geopolitical support for cyber attacks lies a new idea of assault carried out by Dark Team Strom, which has become a personal motive referring to economic (business) gain, where the success of hacking against Israel and pro-Ukrainian countries is used as advertising to attract anyone who wishes to utilize the services of Dark Strom Team.

Date	hacker names	Origin hacker	Attack Target	Impact of Attack
06/10/23	Cyber Av3ngers (Hactivist)	Iranian IRGC Cyber Persona	The Noga Independent Systems Operator (Israeli electric system management)	hacking and launching a Distributed Denial of Service (DDoS) attack.
07/10/23	Anonymous Sudan	Pro Hamas tied to Russia	emergency warning systems (The Alert Applications)	launched DDOS attacks on all the alert applications used for notifying citizens about incoming rockets.
10/10/23			The Jerusalem Post,the largest English-language daily newspaper	The attack on The Jerusalem Post website, caused it to be offline for more than 2 days.
08/10/23	Killnet	Hacker Pro Russia	Israeli government websites.	Announced has launched attacks against Israeli government websites.
19/10/23	Ghosts of Palestine	A Palestinian Hacker Gang	Attack private and public infrastructure in Israel and the United States	Invited hackers from around the world to attack private and public infrastructure in Israel and the United States.
09/10/23	Libyan Ghosts	Libyan Cyber Army	Small Israel Websites	Libyan Ghosts has started defacing small Israeli websites in support of Hamas.
15/11/23	Storm -1133	Groups work for Hamas	Israeli organizations in the defense, energy and telecommunications sectors	Storm-1133 combines social engineering and fake profiles on LinkedIn disguised as human resource managers, project coordinators, and software developers from Israel. sending phishing messages, conducting reconnaissance, and distributing malware.
14/10/23	Cyber Av3ngers	Iranian Group Pro Palestine	- ORPAK (a company provides payment and management solutions for fuel, retail, & fleet businesses in Israel) - CCTV Footage and data from multiple gas stations	Announce that they have compromised ORPAK, leaking CCTV footage and data from multiple gas stations and screenshots of the internal panels using SiteOmat.
10/10/23	Blackfield	A Treat Actor which target United States and Israel	Targeted attacks and disinformation campaigns.	announces on a Russian-speaking forum that they possess data belonging to hundreds of IDF soldiers and Shabak members, including phone numbers, photos, and personal information.
15/10/23	AnonGhost	Hacktivist group from Indonesia	Dating and Consolidation Project for Israeli LGBTQ Community	leak the database of a dating and consolidation project for the Israeli LGBTQ community
16/10/23	YourAnon T13x		Israli news websites "All IsraelNews & Abu Ali Express"	DDoS attacked to the websites.
Hackers who are Pro against Israel				
09/10/23	Lahav 433 Unit & Binance	The Cyber Branch of the Israel Police's	Hamas's Cryptocurrency accounts	freezes cryptocurrency accounts belonging to Hamas.
09/10/23	Garuna & TeamHDP	**Pro Israel**	Websites of Hamas& Islamic University of Gaza	has targeted the websites of Hamas and the Islamic University of Gaza.
10/10/23	ThreatSec	**Pro Israel Group**	Infrastructure of Gaza - based ISP AlfaNet	claims to have compromised the infrastructure of Gaza-based ISP AlfaNet.

Fig. 3. Attacks on Israel, Palestine & Impact of Attacks

4.11 Solomon's Ring

Solomon's Ring is a hacktivist group that is pro-Palestine. Hacktivists on October 7, 2023, claimed to have stolen data from important data centers in Israel. However, the claim regarding the data theft that was to be disseminated by Solomon King remains unproven, as the hacktivists did not send it to the Telegram group.

4.12 Indian Cyber Force

The Indian Cyber Force is a pro-Israel hacker group that supports Israel by targeting cyber attacks against Palestine, including disabling the websites of the Palestinian National Bank and Hamas on October 8, 2023, with both sites being inaccessible for two days.

In addition to the hacktivists mentioned above, there are also several hackers who openly express their support for Palestine and Israel, such as ThreatSec, a pro-Israel group that claimed responsibility for hacking the ISP infrastructure of AlfaNet located in Gaza. The venerable Turkish hacking collective, AslanNeferler Tim, asserts that it executed assaults against an Israeli weaponry producer on October 19 and the Israeli Air Force on October 15. Furthermore, some hacktivists who support pro-Israel causes, like Garuna and TeamHDP, have targeted Hamas websites and the Islamic University of Gaza. The cyber attacks against Israel amid the conflict with Palestine appear quite complex, not only in terms of the number of attackers, both hacktivists and hackers claiming affiliation with other countries, targeting Israel and Palestine. Amidst the numerous DDoS cyberattacks targeting the government and critical infrastructure of Israel, there are many hackers who identify themselves as hacktivists with various motives, making it quite challenging to identify the perpetrators of these attacks. However, claims regarding

the attacks are difficult to identify, especially since most of the hacktivists make their claims through Telegram groups followed by hundreds of thousands of members.

5 Cyber Walfare Landscape

With the advancement of information technology, cyber attacks have become one of the main threats in the context of modern geopolitics. These attacks not only target a country's vital infrastructure but also have deeper motives, aiming to influence interstate relations and global geopolitical interests. In this context, cyber attacks are not just a threat to a country's security but also an instrument used to support political agendas and the interests of other countries.

Cyber attacks against Israel have mostly originated from outside Palestine since the Israeli government cut off internet access for millions of Palestinian citizens. Hacker groups are divided into two categories: attackers supporting Palestine and attackers supporting Israel. The attack methods used by hackers involved in cyber warfare include Distributed Denial of Service (DDoS) attacks, malware distribution, phishing campaigns, and website defacement.

The cyber attacks in Israel highlight the political complexity in the Middle East, where factors such as political tension, religion, and ideology often trigger conflicts. As we know, cyber attacks are not just an attempt to damage digital infrastructure but also a form of political support for countries or groups involved in regional conflicts. Therefore, as additional data in this research, the researcher has included some information regarding the cooperation between Israel and Palestine in the field of cybersecurity, both in strengthening offensive cybersecurity as well as in defensive security and other collaborations in the cyber domain.

5.1 Nations Engaged in Cyber Warfare for Palestine

What is quite interesting in the landscape of war between Israel and Palestine is not only the response and support of states in the real world but also in the cyber sphere both countries are getting support from various countries. The country seeks to provide support for Palestinian cybersecurity both in response to cyber attacks and in preventing cyberattacks offensively and defensively.

5.1.1 Iran

Iran is one of the countries that strongly supports Palestinian independence both in the ongoing real-world war between Palestine and Israel and in the cyber war between Israel and Palestine. Iran's support for Palestine in the cyber war is using its remarkable expertise and technological capabilities in various forms of cyber attacks such as Distributed Denial of Service (DDoS), ransomware attacks, and infiltration of enemy networks. Iran also plays an important role in protecting Palestinian cyber infrastructure and preventing Israeli cyberattacks.

Iran has also contributed to the cyber war by launching Israeli technological infrastructure attacks with the aim of interfering with Israeli technology systems. These attacks

are aimed at suppressing and technologically damaging Israel, thus making it easier for Iran to support Palestine in its cyber conflict with Israel. In addition, Iran provides technological and cyber security support to Palestine by helping Palestine respond to Israeli cyber attacks in a cyber-defensive manner. Iran's cyber defence is carried out by providing the technology, cyber expertise, and cybersecurity training needed by Palestinian fighters increases their capacity to prevent and counter Israeli attacks.

In October 2023, Iran intensified its influence operations in support of Hamas during the Israel-Hamas conflict. These operations combined cyber attacks with manipulative messages on social media to influence perceptions and behaviors. Iran utilized fake online accounts, or "sockpuppets," to disseminate anti-Israel messages, intimidate Israeli citizens, and criticize the Israeli government, aiming to escalate regional tensions. Iran also targeted Israel's allies and economic partners with cyber attacks and influence campaigns to weaken support for Israel. The threat from these operations is expected to grow, particularly leading up to the US elections in November 2024.

Iran's operations have primary goals: retaliation, intimidation, and reducing international support for Israel.

a. Retaliation

   Iran uses cyber operations to retaliate against Israeli actions, including disrupting the supply of electricity, food, water, and fuel to Gaza City, as well as targeting Israel's infrastructure in these areas.

b. Intimidation

   Iran's operations are designed to intimidate Israelis and their supporters through threatening messages, creating a perception that their infrastructure and governance systems are vulnerable.

c. Motive to weaken international support for Israel

   Iran has a goal of reducing state support for Israel by using its support for both political, war and cyber warfare. Iran also highlights the damage caused by Israeli attacks on Gaza by pointing out that Israel's actions are genocidal war crimes.

5.1.2 Türkiye

Turkey is also a country that has a foreign policy in supporting Palestine both politically and militarily in the Israeli-Palestinian conflict. Turkey's support for Palestinian cyber security includes financial assistance to strengthen its technological infrastructure. Turkey also offers cyber training and education to Palestinian cyber experts with the aim of enhancing Palestinian capabilities in response to defensive cyber attacks by Israel. In addition to financial assistance and cyber-training to Palestine, Turkey is also active in developing technological infrastructure in Palestine with the help of the establishment of a strong cyber security system and adequate telecommunications infrastructure. Through these efforts, Turkey has made a real contribution to strengthening Palestinian cyber security in dealing with cyber attacks from both Israel and other countries.

In the Israeli-Palestinian war, Turkey also took part in the cyber war by using various tactics such as launching DDoS attacks on Israeli websites. In addition, Turkish hackers are also involved in the leakage of Israeli government data, disclosing classified information to the public in order to reveal unethical practices or violations by the Israeli Government against the Palestinians. Turkish hackers are also engaged in disseminating

pro-Palestinian messages through social media, with the aim of informing Palestinians in the midst of the Israeli and Palestinian wars and gaining support for Palestine. Turkish hackers are spreading information that seeks to raise global awareness of the world's support for Palestine. In the cyber welfare Turkey took part by retaliating against Israeli cyber attacks against Palestinians. In addition, Turkey provides cybersecurity protection services to Palestinians to safeguard their vital systems.

5.1.3 Qatar

Qatar is a country that has played an active role in supporting Palestine in the cyber war against Israel, by providing vital protection in Palestinian cybersecurity amid military conflict and cyber conflict. Qatar's assistance in Palestine cyber security is a financial aid used by Palestine to strengthen its technological infrastructure to prevent cyber attacks. In addition to financial assistance, Qatar is also seeking to offer training and education for Palestinian cyber experts so that Palestine can cope with an effective Israeli cyber attack.

5.1.4 Malaysia

Palestinian cybersecurity support also supports Malaysia by extending support to Palestinians in the cyber war against Israel by providing adequate financial assistance, training, and technological infrastructure. Like other countries, Malaysia also provides financial assistance and training to Palestinian cyber experts with a view to preventing cyber attacks. Malaysia's role in providing technology infrastructure to Palestine is to counter Israel's cyberattacks on Palestine.

5.1.5 Indonesia

Indonesia is a very active country speaking about Palestinian independence and security in the international arena. In addition to domestic and foreign political support for Palestine, Indonesia is also actively supporting Palestine in the cyber war against Israel. The Indonesian Government is also making every effort to build partnerships and collaborations with international nations that are useful in expressing support against cyber attacks directed at Palestine. Indonesia has allocated funding to Palestinians to enhance cyber capabilities, strengthen infrastructure and help with cyber security training. Moreover, Indonesia has implemented measures to protect Palestine from cyber attacks, as demonstrated by the launch of counter-attacks against Israel. Furthermore, Indonesia is collaborating with other countries and international bodies to improve Palestinian cybersecurity. Indonesia's dedication to resolving peace and justice for the Palestinian people is demonstrated in its efforts to support Palestine in the cyber war against Israel.

5.2 The Role of Israel's Cyber Warfare Allies

The participation of other nations in the Israeli-Palestinian conflict in the domain of cyberspace is a customary aspect of cyber warfare. If there is an ongoing military conflict between two governments that involves the participation of other nations, a cyber war

presents a significant chance for additional countries to align with their allies. The Israeli allies actively participate in the cyber warfare against Israel by targeting and sabotaging its technological infrastructure, disseminating harmful propaganda against Palestine, and seeking to breach the security of Palestinian infrastructure, so endangering the data and information security of Palestinians. Curiously, the nations that openly endorse the autonomy of Israel also engage in cyber warfare against Palestine. These nations include the United States, Canada, Australia, and Germany. These countries possess cybersecurity capabilities and employ highly advanced technology to execute cyber assaults.

5.2.1 United States

The United States of America is Israel's primary ally in the cyber fight, possessing ample capabilities and resources to carry out cyber operations against perceived cybersecurity threats to Israel. Furthermore, alongside the DDoS attack, the United States employs phishing methods to acquire confidential data from the Palestinian authorities and organizations. The United States has also focused on communications systems that seek to limit access to information and disrupt networks utilized by states supported by Palestinians. Furthermore, the United States provides Israel with technology and cyber education in order to bolster Israel's proficiency and capacity in cyber defense.

5.2.2 Canada

Canada's active participation in bolstering Israel's cyber capabilities is evident through its efforts to destroy the technological infrastructure of a state that supports Palestine. Canada actively supports Israel in the cyber domain by engaging in cyber-interference efforts and limiting the online presence of Palestinian supporters. The objective is to curtail the dissemination of information that is incongruent with Israel. Furthermore, Canada is dedicated to protecting its cyber security by collaborating with Israel in the field of cybersecurity. The relationship between Canada and Israel encompasses the provision of technical support, training, and resources essential for safeguarding Israel against cyber threats originating from Palestinian and pro-Palestinian states.

5.2.3 Australia

Australia is one of the nations that provide support to Israel in the realm of cyber warfare. The Australian Government has forged strong connections with Israel in the realm of cyber security and numerous facets pertaining to cyber defense. Their role involves offering technology support, expertise, and personnel to aid Israel in upholding cybersecurity.

5.2.4 Germany

Through its relationship with Israel, Germany provides the necessary resources and encourages the development of cyber skills and technologies. The German Government is actively aiding Israel in safeguarding its cyber infrastructure against assaults from Palestinian entities. Furthermore, Germany is fostering collaboration with Israel in the

areas of information sharing, training, and strategy creation to uphold cyber security within the country.

5.2.5 India

In fact, Israel and India are recognized as countries with capabilities in technology and cyber security. Therefore, both countries have cooperation in the fields of technology and cyber security based on shared needs and security threats. India and Israel face cyber security threats from both state and non-state actors; thus, Israel and India are strengthening their cooperation in developing cyber defense through cyber training, intelligence sharing, and collaboration in cyber security research. Then, in technological innovation, the two countries collaborate on the development of AI, the Internet of Things (IoT), and application development.

6 Hactivist or Cyber Walfare

The hacking of government websites and various infrastructures in Israel raises the question of whether these attacks were carried out by hacktivists or if they were part of a cyber warfare strategy against Israel. Looking at the various motives of hacktivists, their initial goal seems to be driven by a political agenda, as claimed by the hacktivists themselves. However, over time, with the increasing calls for support for Palestine, some other hacktivists have seized various opportunities in politics, economics, business, and personal beliefs to target Israel.

According to Robert E. Joyce, a Cybersecurity Coordinator on the U.S. National Security Council, on October 9, 2023, there were numerous DDoS attacks, resulting in 60 websites being taken down and more than five sites being compromised. The attacks were found to be predominantly carried out by international hacker groups that are not affiliated with any government, but rather by hackers who identify as hacktivists with various motives, such as being pro-Israel or pro-Palestine, driven by humanitarian, political, and economic reasons. There are also those motivated by advertising (business) for economic gain, showcasing their hacking capabilities against specific websites, such as Israeli government sites. Additionally, many hackers are motivated by personal beliefs or engage in hacking out of their own desires, taking advantage of the existing conflict situation.

Furthermore, it is undeniable that the developments regarding the conflict and the large-scale hacktivist attacks against Israel pave the way for countries like Iran to adopt a hacktivist disguise supported by the Iranian state through massive DDoS campaigns. Essentially, in the Israel-Palestine conflict, the involvement of hacktivists colors the conflict between the two countries; the war is not only supported by states but also by non-state actors who take a stand by transforming warfare into cyber warfare. Furthermore, the presence of hacktivists in the conflict provides an opportunity for other countries to easily enter the fray through the support of hacktivists affiliated with their governments.

In the modern era, characterized by modern warfare such as cyber warfare, there is a significant concern for national security across various sectors connected to the internet. The existence of cyberspace, which can be accessed by anyone, brings both positive and

negative impacts for its users. The escalation of the conflict between Israel and Palestine highlights the dangers of cyber threats, not only from state-to-state cyber warfare but also from groups that are not affiliated with any nation. These groups pose a significant concern as they use all their resources to destroy their opponents through phishing, data misuse, leaking important state information sold on the dark web, hacking critical sectors of the state, both public and private, and damaging systems in both the cyber realm and in tangible forms, all without accountability from anyone.

The presence of hacktivists in the chaos of conflict can exacerbate the situation and the casualties in conflict areas. The motives used by hacktivists vary, such as the humanitarian motive demonstrated by some pro-Palestinian hacktivists, and the geopolitical motive employed by Killnet and Sudan Anonymous, which are affiliated with Russia. Their goal in conducting hacks is to reinforce the message that Russia supports Palestine, and the attacks on Israel are a retaliation for Ukraine's support in becoming part of NATO. The assaults are not only aimed at Israel but also at countries that support Israel, such as United States, Kenya and other nations that have good relations with Ukraine.

The presence of hacktivists in the cyber realm poses a national security threat to countries around the world, including those in conflict, due to the potential invisible hand that may receive support from governments to hack other nations. Even without political backing from other countries, hacktivists will employ various methods with self-serving motives, such as geopolitical support, political agendas, or responses to humanitarian issues, by exploiting the turbulent conditions of conflict between nations.

7 Israel Respond

Israel's strength in the field of cybersecurity is undeniable, as it has become a world leader in cybersecurity. Israel is home to 450 cybersecurity startup companies with a sales value of $7.5 billion in 2023. The awareness of the importance of the internet and the threats of cyber security has been recognized by Israel since the early spread of the internet. Therefore, Israel has established Unit 8200 as part of the Israel Defense Forces to develop technology, which is known as the most advanced intelligence defense unit in the world.

Israel is one of the countries that has a well-funded environment for cybersecurity and cyber defense. As a country facing geopolitical threats, including conflicts with Palestine and other countries such as Iran, Lebanon, and other terrorist groups that are not affiliated with any states, Israel is strengthening its offensive and defensive security capabilities in the military and cybersecurity fields.

Israel has a history of various past events that have led the Israeli government to focus on national security not only on military security but also on cyber security in facing cyber conflicts in the modern era. Interestingly, various events occurring both domestically and on the international stage have led Israel to learn a lot in order to strengthen its cyber security. The following is one of the "Israeli Cyber Defense Reports" as part of Israel's ongoing efforts to develop its cyber security, conducted by the Israeli National Cyber Security and Cyber Defense Posture.

The development of Israel's cyber security is largely based on cyber attack events that occur within Israel as well as international incidents that shape Israel's cyber security

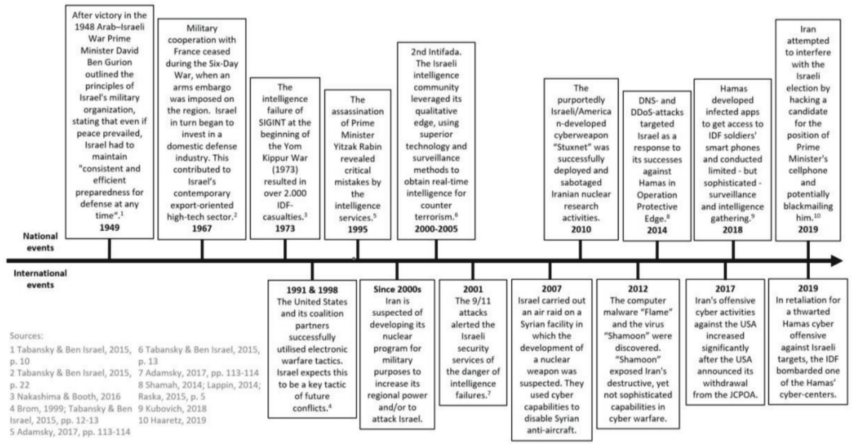


Fig. 4. The history of cyber attacks shapes Israel's cybersecurity strategy

policy. Below is a "cyber defense report" by the Cyber Defense Project (CDP), Center for Security Studies (CSS), and ETH Zürich regarding the development of Israel's cybersecurity policy over time, based on events that occurred (Fig. 4) both internally within the country and internationally outside of Israel (Fig. 5).

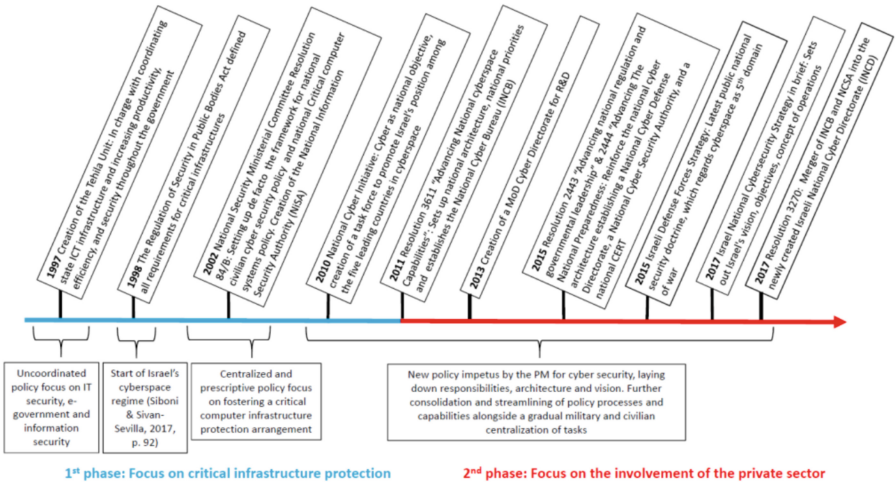


Fig. 5. The evolution of Israel's cyber policy over time.

Based on the incident reports that underlie Israel's cybersecurity policy, it can be seen that the Israeli government has made various efforts to maintain national security in the face of modern warfare challenges such as cyber warfare. In March 2019, Israel announced its readiness for an information war with Palestine, considering that it is a strong country in cybersecurity, as Israel understands the importance of cybersecurity in

the ongoing conflict in the Middle East. On the other hand, Palestinian hackers sometimes succeed in breaching Israel's data security leaks. Israel also uses Unit 8200 to conduct surveillance on Palestinians through the interception of communication media such as emails, phone calls, and Palestinian social media as one of the tools to divide Palestinians.

However the cyberattack incident that occurred in Israel, carried out by hacktivists, highlights to the world that no security system is truly strong and secure. As a leading cyber nation, Israel failed to detect the attacks, both from Hamas through online intelligence and satellites, as well as the DDoS attacks against the Israeli government conducted by pro-Palestinian hacktivists. The failure to detect and counter the attacks from both Hamas and the hacktivists demonstrates to the world that Israel still has vulnerabilities in its surveillance and monitoring systems in the online space.

The success of cyberattacks by hacktivists and some hackers affiliated with pro-Palestinian states has opened the eyes of the international community to the importance of cybersecurity in various critical infrastructures within countries. Israel, as a country with capabilities in the field of cybersecurity and recognized as a global cyber leader, is still caught off guard in facing international hackers. The limitations in facing hacktivists due to the lack of Israel's cyber defense preparedness stem from systemic issues, including data security vulnerabilities due to negligence, inadequate defense system planning, weak passwords, and the failure to regularly update operating systems and software. In addition, the large number of medium-sized companies affected by cyber attacks due to the high costs of red teaming security testing has led these companies to neglect identifying weaknesses in their security systems. As a result, the rapid process of digitalization has increased Israel's vulnerability to cyber attacks targeting critical infrastructure, while also facilitating criminal activities such as identity theft and credit card fraud.

Overwhelmed in facing cyber hacks targeting many sites, both government, public, and private. Israel has experienced a significant increase in cyberattacks from hacktivist groups, especially during intense conflict periods such as the Israel-Hamas war. These attacks have exploited vulnerabilities in Israel's cybersecurity, causing disruptions to various digital systems and services. The impact of cyber attacks on ICS and SCADA systems can lead to dangerous disruptions in vital infrastructure services such as electricity, water, communication, and transportation. This disturbance is a major concern in terms of public safety, national security, and it also impacts the economy.

According to data from the Israeli National Cyber Directorate dated May 9, 2023, the country's economy suffered an annual loss of 12 billion Shekels ($3.3 billion) due to cyberattacks. These losses include short-term costs such as payments to the initial response team, direct financial impacts like downtime, data recovery, and legal advice, as well as medium to long-term costs including lost future revenue, missed business opportunities, reputational damage, lawsuits, and third-party losses. This analysis considers various types of attacks that cause business losses for organizations using both "top-down" and "bottom-up" approaches. The chief economist at the National Cyber System stated, "This data reveals for the first time the significant economic burden caused by cyber attacks on national products and the cumulative losses to the economy." Without steps to enhance cybersecurity at both the individual and organizational levels,

these costs will continue to rise each year as technological advancements persist across all areas of life.

Israel has responded to cyber attacks with a diverse approach, leveraging its advanced cybersecurity capabilities and strategic partnerships. The country has developed a robust cyber defense strategy, focusing on early detection, rapid response, and resilience-building measures. Israel and the US have made a cybersecurity collaboration agreement to combat terrorist financing, specifically focusing on cyber research and collaboration on transportation cybersecurity. The goal of the cooperation between the US and Israel in the cyber field is to protect the economies and critical infrastructure of both countries, as well as to enhance collaboration in managing the response to cyber threat risks.

Israel's efforts to advance its cyber security capabilities and security are indeed remarkable. As knowledge and technology rapidly advance, the challenges and threats of cyber security that all countries around the world will face will also evolve significantly. Therefore, the importance of an early approach to cyber security for the international community is crucial, not only for Israel and countries aligned with Israel in terms of cyber security but also for all nations that consider cyber security to be vital for their safety. Researchers, on this occasion, provide several policy recommendations that can be used to further strengthen national and international cybersecurity as follows:

7.1 Inclusive Cyber Security Education for All Levels

One of the obstacles in implementing cyber security in the country is the general public's lack of knowledge about security and cyber threats. Therefore, the government must implement a cyber education curriculum that is not only focused on universities but also on elementary and secondary schools so that they can recognize what cyber threats look like.

7.2 Raising Public Awareness About Cyber Security

Cyber security campaigns are also important to promote not only through research but also through promotion on social media, television, and other platforms that can inform the public about cyber security.This makes all members of society recognize the latest cyber threats so they can be more cautious.

7.3 Collaboration

Between the internal NGO, INGO, and Government, as well as international collaboration, to combat cybercrime.

7.4 Ensuring Continuous System Updates for All Devices

It is essential to ensure that all devices, including phones, tablets, and laptops, are always updated. Automatic updates for operating systems and security software should be enforced, and users need to be educated about the importance of routinely updating their devices to prevent vulnerabilities that can be exploited by cybercriminals.

7.5 Securing Critical Infrastructure, Public and Private

Securing critical infrastructure is very important. Systems that manage energy, water, transportation, and communication must be strictly protected. The public and private sectors must collaborate to create a robust security system to ensure that data is protected across industries such as finance, healthcare, and telecommunications. Applying the "Zero Trust" security model ensures that every access must be verified first.

7.6 Conducting Active Security Assessments

Routine security checks are necessary to identify and rectify weaknesses in IT systems. This includes penetration testing, which simulates cyberattacks to identify security vulnerabilities. Continuous monitoring of endpoints, applications, servers, and networks will also help detect potential threats earlier. Regular incident response drills will ensure that the systems and teams are prepared to respond effectively in the event of an attack.

8 Conclusion

Based on this research, it can be seen whether the attacks carried out by international hackers against Israel are part of hacktivism or cyber warfare against Israel in support of Palestine. The results of the research indicate that the hacking attacks on critical infrastructure, government websites, and the "Red Alert" rocket announcement application to the Israeli public are actions carried out by international hacktivists. However, it cannot be denied that the presence of hacktivists in the Israel-Palestine conflict provides an open opportunity for Israel's enemy states, such as Iran, to utilize the term hacktivist or hacktivist services to launch cyberattacks against Israel. The motivations behind various hacktivist attacks on Israel are diverse; they are not only rooted in political and religious support for Palestine but also based on economic motives, such as selling services for attacks against Israel. Additionally, the motivations for attacking Israel are also grounded in the self-beliefs of the hacktivists. Then Israel, as a leading cyber nation in the world, has taken several steps to prevent attacks from hacktivists. Israel has responded to cyber attacks with a varied approach, leveraging its advanced cyber security capabilities and strategic partnerships, such as with the United States. The country has developed a robust cyber defense strategy, with measures to build resilience, emphasizing early detection and rapid response.

References

1. Tabandky, L.: How Israel Became a Top Cyber Power, 21 Mar 2022.https://nationalinterest.org/blog/techland-when-great-power-competition-meets-digital-world/how-israel-became-top-cyber-power. Accessed 9 Feb 2024
2. The Hindu Bureau: Hamas - Israel Conflict enters Cyber Space as Hacker Groups Target Government and Private Digital Assets: Report, 10 October 2023. https://www.thehindu.com/sci-tech/technology/hamas-israel-conflict-enters-cyber-space-hacker-target-government-private-assets/article67403631.ece. Accessed 9 Feb 2024

3. Kate, L.: 'Hacktivism', Cyber Warfare, and the 2023 Israel - Hamas Conflict, 25 November 2023. https://www.theyoungdiplomats.com/post/hacktivism-cyber-warfare-and-the-2023-israel-hamas-conflict. Accessed 7 Feb 2024
4. Andika, R.: Usai Nyatakan Perang, Lebih dari 60 Website Israel Lumpuh Akibat Serangan Hacker, 10 October 2023. https://fourtrezz.co.id/usai-nyatakan-perang-lebih-dari-60-website-israel-lumpuh-akibat-serangan-hacker/. Accessed 7 Feb 2024
5. Microsoft: Microsoft Digital Defense Report 2023 Building and Improving Cyber Resilience, October 2023. https://www.microsoft.com/content/dam/microsoft/final/en-us/microsoft-brand/documents/MDDR_FINAL_2023_1004.pdf. Accessed 8 Feb 2024
6. Cyfirma Decoding Threats: Israel Gaza Conflict: The Cyber Perspective, 18 October 2023. https://www.cyfirma.com/outofband/israel-gaza-conflict-the-cyber-perspective/. Accessed 8 Feb 2024
7. Tal, L.: Israel - Hamas 2023 Symposium - Cyberspace - The Hidden Aspect of The Conflict, 30 November 2023. https://lieber.westpoint.edu/cyberspace-hidden-aspect-conflict/. Accessed 16 Feb 2024
8. Marco, R., Niek Jan, H.D.V.: Hacktivism and website defacement: motivations, capabilities and potential threats. In: Virus Bullettin Conference October 2017. Netherlands: The Hague University to Applied Sciences (2017)
9. Mohammad, L.S.A., Yildirim, T.: Israeli - Palestinian cyber conflict. Eskişehir Osmangazi Üniversitesi İİBF Dergisi **16**(1), 186–204 (2021). https://doi.org/10.17153/oguiibf.869178
10. Beatrice, G.: From Kinetic to Cyber Attacks and Back: The Israeli approach to Deterrence in Cyberspace and The Multi - Dimensional Threat of Hamas. Universita Degli Studi Firenze: Center for Cyber Security and International Relations Studies, February 2022. https://www.cssii.unifi.it/Is-6-cyber-security.html
11. Fabio, C.: Deterritorializing Cyber Security and Warfare in Palestine: Hackers, Sovereignty and the National Cyberspace as Normative. Leiden University, Netherlands. Cyberorient, vol. 13, no. 1, pp. 28–42 (2019). https://ssrn.com/abstract=3640789
12. Sarah, B.: Aninymous Sudan Targets Israel's Critical Infrastructure, 02 November 2023. Accesses from University of Hawai'I - West O'AHU: https://westoahu.hawaii.edu/cyber/global-weekly-exec-summary/anonymous-sudan-targets-israels-critical-infrastructure/. Accessed 19 Feb 2024
13. Christopher, C., Marcin, N.: Anonymous Sudan Campaign Analysis, 06 November 2023. Accesses from NETSCOUT: https://www.netscout.com/blog/asert/anonymoussudan#:~:text=Anonymous%20Sudan%20emerged%20in%20a,%2D%20or%20Sudan%2Drelated%20matters. Accessed 19 Feb 2024
14. Cloudflare: What is Anonymous Sudan (2023). https://www.cloudflare.com/en-gb/learning/ddos/glossary/anonymous-sudan/. Accessed 19 Feb 2024
15. Reichman University, International Institute for Counter - Terrorism (ICT) Anonymous Sudan: Pro - Islamic Hacker Group Engages in Cryptocurrency Donation Campaign, 27 November 2023. https://ict.org.il/anonymous-sudan-hacker-group/. Accessed 19 Feb 2024
16. Flashpoint Intel Team: Unmasking Annymous Sudan: Timeline of DDoS Attacks, Affiliations and Motivations, 20 June 2023. https://flashpoint.io/blog/anonymous-sudan-ddos-timeline/. Accessed 20 Feb 2024
17. Kumar, R.R.: DDoS Attacks and The Cyber Threatscape, 01 August 2023. Accesses from MP - IDSA Issue Brief: https://www.idsa.in/system/files/issuebrief/ib-DDoS-Attacks-rksharma-020823.pdf. Accessed 20 Feb 2024
18. ThreatMon, S.: Anonymous Sudan: in - Depth Analysis Beyond Hacktivist Attacks (2023). https://threatmon.io/storage/anonymous-sudan-in-depth-analysis-beyond-hacktivistattacks.pdf. Accessed 20 Feb 2024

19. Claire, R.: Iranian Cyber Criminals Targetiing Israel Technology Hack Into Pennyslvania Water System, 29 November 2023. https://www.usatoday.com/story/news/nation/2023/11/29/irans-cyber-av3ngers-target-israel-unitronics-devices-in-us/71741103007/. Accessed 21 Feb 2024
20. Radware: Cyber Aggression Rises Following the October 2023 Israel - Hamas Conflict, 11 October 2023. https://www.radware.com/security/threat-advisories-and-attack-reports/cyber-aggression-rises-following-the-october-2023-israel-hamas-conflict/. Accessed 22 Feb 2024
21. Antoaneta, R., Maggie, M.: How hackers piled onto the Israeli - Hamas Conflict, 15 October 2023. https://www.politico.eu/article/israel-hamas-war-hackers-cyberattacks/. Accessed 23 Feb 2024
22. Aqil, H.: Nearly 100 Hacker Groups Take Israel - Hamas Conflict Into Cyberspace By Waging Online Proxy War, 12 October 2023. https://www.straitstimes.com/world/hackers-take-sides-in-israel-hamas-conflict-by-waging-proxy-war-in-cyberspace. Accessed 26 Feb 2024
23. Security Scorecard: Cyber Risk Intelligence Update: Hacktivist Involvement in Israel - Hamas War Reflects Possible Shift in Threat Actor Focus (2023). https://securityscorecard.com/research/hacktivist-involvement-in-israel-hamas-war-reflects-possible-shift-in-threat-actor-focus/. Accessed 28 Feb 2024
24. Patrick, L.: Code of Conflict: The Global Cyber Divide Between Gaza and Israel (2023). https://www.securityhq.com/blog/code-of-conflict-the-global-cyber-divide-between-gaza-and-israel/. Accessed 29 Feb 2024
25. Abdul, R.H.: Israel Hamas War: Indian Cyber Space Caught in the Crossfore As Threat Groups Escalate Online Conflict, 20 October 2023. https://business.outlookindia.com/news/israel-hamas-war-indian-cyber-space-caught-in-the-crossfire-as-threat-groups-escalate-online-conflict. Accessed 29 Feb 2024
26. Paulina: Bangladesh Hacktivists Target Critical Infrastructure in India, Israel, and Australia, 15 November 2023. https://cybernews.com/news/mysterious-team-bangladesh-hacktivists-attack-india-israel-australia/. Accessed 1 Mar 2024
27. Ashish: Pro - Palestine Hackers Target Israeli Security Agency, The Occupation Magazine (2023). https://thecyberexpress.com/cyberattacks-on-israel-israel-palestine-war/amp/. Accessed 4 Mar 2024
28. K. Cyber Defense Insight: Indonesian Hackers Launch Cyber Attacks on Israel Amidst Hamas Surge, 09 October 2023. https://www.cyberdefenseinsight.com/2023/10/indonesian-hacker-launch-cyber-attacks.html. Accessed 7 Mar 2024
29. Lily Hay, L.N., Matt, B.: Activist Hackers Are Racing Into The Israel - Hamas War - for Both Sides, 09 October 2023. Accesses from wired security: https://www.wired.com/story/israel-hamas-war-hacktivism/. Accessed 7 Mar 2024
30. Ashish, K.: Cyber Walfare Surges as Over 35 Hacktivist Groups Join Israel - Palestine Conflict. The Cyber Assaults, Predominantly Distributed Denial of Service (DDoS) attacks, Have Been Orchestrated Against Critical Israeli Infrastructure, 11 October 2023. Accesses from Firewall Daily, Hacker News: https://thecyberexpress.com/israel-palestine-conflict-cyber-warfare-risk/. Accessed 7 Mar 2024
31. Daniel, S.: Hacktivists Bring OpIsrael Back to Life, 14 October 2023. https://www.radware.com/blog/security/threat-intelligence/2023/04/hacktivists-bring-opisrael-back-to-life/. Accessed 7 Mar 2024
32. Cyber Threat Landscape: #OpIsrael, #FreePalestine and #OpSaudiArabia - How Cyber - Threat Actors Coordinate PSYOPS Campaigns with Kinetic Military Actions, 09 October 2023. Accesses from Resecurity: https://www.resecurity.com/blog/article/opisrael-freepalestine-and-opsaudiarabia-how-cyber-actors-capitalize-on-war-actions-via-psy-ops. Accessed 7 Mar 2024

33. Microsoft Security: Iran Suges Cyber - Enabled Influence Operations in Support of Hamas 26 February 2024. Accesses from Iran surges cyber-enabled influence operations in support of Hamas | Security Insider (microsoft.com). Accessed 5 Mar 2024
34. Sam: Iranian Cyber Capabilities: Assessing the Threat to Israeli Financial and Security Interests. Cyber, Intelligence and Security, vol. 3, no. 1, May 2019. Accesses from Cyber3.1ENG_3-73-96.pdf (inss.org.il). Accessed 5 Mar 2024
35. Ashmita, S.: Cyber Walfare Escalates as Hackers Target Websites Amid Israel - Hamas Conflict, 10 October 2023. Accesses from India Today: https://www.indiatoday.in/world/story/cyber-warfare-escalates-as-hackers-target-websites-amid-israel-hamas-conflict-2447045-23-10-10. Accessed 9 Mar 2024
36. Coker, J.: Growing Concern Over ROle of Hacktivism in Israel Hamas Conflict, 16 October 2023. Accesses from Infosecurity Magazine: https://www.infosecurity-magazine.com/news/concern-hacktivism-israel-hamas/. Accessed 9 May 2024
37. Yoachimik, O., Pacheco, J.: Cyber Attack in The Israel - Hamas War, 23 October 2023. Accesses from Cloudflare: https://blog.cloudflare.com/cyber-attacks-in-the-israel-hamas-war/. Accessed 10 May 2024
38. Altshuler, T.S.: Israel's Cybersecurity is a Tickng Time Bomb - Opinions, 05 January 2023. Accesses from The Jerusalem Post: https://www.jpost.com/opinion/article-726659. Accessed 11 June 2024
39. Jeffay, K.: Israel Was The Number One Cyber Attack Target in 2023, 10 April 2023. Accesses from Israel21c: https://www.israel21c.org/israel-was-the-number-one-cyber-attack-target-in-2023/. Accessed 11 June 2024
40. Ani: Israel Saw 43 Pc Increase in Reported Cyber Attacks in 2023. Accesses from South Asia Leading Multimedia News Agency, 01 April 2024. https://www.aninews.in/news/world/middle-east/israel-saw-43-pc-increase-in-reported-cyber-attacks-in-2023202404012239 31/. Accessed 11 June 2024
41. Ani: Israel's Economy Loses Billions a Year due to Cyber Attacks, 09 May 2024. https://theprint.in/world/israels-economy-loses-billions-a-year-due-to-cyber-attacks/2077990/. Accessed 10 May 2024
42. Daniel, L.: The Territorialization of Cyberspace. International Studies Review, undefined (2020). https://doi.org/10.1093/ISR/VIZ022. Accessed 25 Sept 2024
43. Gallagher, R., Robertson J.: Cyberattacks Targeting Israel Are Rising After Hamas Assault, 10 October 2023. https://time.com/6322175/israel-hamas-cyberattacks-hackers/. Accessed 22 Sept 2024
44. Security Scorecard: Cyber Risk Intelligence Update: Hacktivist Involvement in Israel - Hamas War Reflects Possible Shift in Threat Actor Focus (2023). https://securityscorecard.com/research/hacktivist-involvement-in-israel-hamas-war-reflects-possible-shift-in-threat-actor-focus/. Accessed 22 Sept 2024
45. Taddeo, M.: The Limits of Deterrence Theory in Cyberspace, vol. 31, pp. 339–355 (2017). https://link.springer.com/article/10.1007/s13347-017-0290-2
46. Kovacs, E.: Hackers Join in on Israel - Hamas War with Disruptive Cyberattacks, 09 October 2023. https://www.securityweek.com/hackers-join-in-on-israel-hamas-war-with-disruptive-cyberattacks/. Accessed 22 Sept 2024
47. Hendriyanti, M.R.: Hacktivist Hackers Disable Israeli Websites Impact of Hamas Vs Israel War?, 13 January 2024. https://www.liputan6.com/tekno/read/5419321/hacker-hacktivist-lumpuhkan-situs-situs-israel-imbas-perang-hamas-vs-israel?page=4. Accessed 22 Sept 2024
48. Frei, J.: Cyberdefense Report, Israel's National Cybersecurity and Cyberdefense Posture. Cyber Defense Project (CDP). Center for Security Studies (CSS). ETH Zurich (2020). https://doi.org/10.3929/ethz-b-000438397

49. Startup Nation Central.: Cybersecurity in Israel: Leading the Words to a Safer Digital Future, 06 March 2024. https://startupnationcentral.org/blog/cybersecurity/cybersecurity-in-israel/. Accessed 29 Sept 2024
50. Langley, K.: Hacktivism, Cyber Walfare and the 2023 Israel - Hamas Conflict, 25 November 2023. Accesses from Young Diplomats Society: https://www.theyoungdiplomats.com/post/hacktivism-cyber-warfare-and-the-2023-israel-hamas-conflict#:~:text=A%20recent%20intelligence%20update%20by%20a. Accessed 29 Sept 2024
51. Cyber Security Intelligence: The Israeli - Hamas Conflict Shows Cyber Walfare is the New Normal, 17 October 2023. https://www.cybersecurityintelligence.com/blog/the-israeli-hamas-conflict-shows-cyber-warfare-is-now-the-new-normal-7250.html#:~:text=Cyber%20attacks%20in%20the%20Israel-Palestinian. Accessed 30 Sept 2024

Designing a Secure E-Voting System with Ganache: Challenges and Solutions

Garima Mathur(✉), Pushpraj Singh Chauhan, and Privi Dubey

Sagar Institute of Science and Technology, Bhopal, Madhya Pradesh, India
{garimamathur,prividubey}@sistec.ac.in

Abstract. A fair and secure voting mechanism is essential for democratic systems to function effectively. Current electronic voting systems, particularly those relying on Electronic Voting Machines (EVMs), exhibit vulnerabilities such as susceptibility to hacking, data tampering, and lack of verifiability. This paper proposes a secure electronic voting system leveraging the Ethereum blockchain and smart contracts to mitigate these challenges. Our system introduces decentralized voting processes with cryptographic mechanisms to ensure transparency, voter anonymity, and resistance to fraud. We address issues like double voting and vote tampering by utilizing blockchain's immutability and distributed ledger technology. The proposed solution also includes mechanisms for individual verifiability, where voters can confirm the accuracy of their vote, and universal verifiability for the integrity of the overall results. The system uses Ganache for blockchain simulation, with smart contracts managing the election process. Transaction efficiency and network congestion are handled using optimized smart contract designs. Future work includes improving scalability and adopting advanced consensus algorithms to enhance security and transaction throughput.

Keywords: Blockchain voting · Ethereum smart contracts · cryptographic security · voter anonymity · verifiability

1 Introduction

As the global shift from traditional paper-based voting to electronic voting systems accelerates, ensuring the security, transparency, and integrity of the voting process has become a critical challenge. Current electronic voting systems, particularly those using Electronic Voting Machines (EVMs), are vulnerable to a variety of security threats, including unauthorized access, data tampering, and system hacking. These machines often store data locally and centrally, making them susceptible to manipulation, as demonstrated by cases of "booth hijacking" and other forms of election fraud [1, 2]. To address these vulnerabilities, this paper introduces a blockchain-based electronic voting system using Ethereum and smart contract technologies. Unlike traditional centralized systems, blockchain's decentralized and distributed ledger architecture provides a more secure and transparent voting mechanism. By leveraging cryptographic hashes and the immutability of blockchain records, this approach ensures that once a vote is cast, it cannot be altered, thereby preventing fraud and tampering [5, 6].

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 233–244, 2025.
https://doi.org/10.1007/978-3-031-86305-9_16

Ethereum smart contracts play a key role in this system by automating critical processes such as voter authentication, ballot management, and vote tallying. These contracts are self-executing scripts that run on the Ethereum Virtual Machine (EVM), ensuring the enforcement of predefined rules without relying on intermediaries. This decentralized execution ensures that the election process is resistant to external attacks or internal manipulation. Moreover, our system employs cryptographic mechanisms to guarantee voter anonymity, using pseudonyms as voter addresses while maintaining the verifiability of individual votes. This ensures that voters can verify their votes were accurately recorded without revealing their identity. The system also supports universal verifiability, allowing third parties to audit the election outcome and confirm its integrity. By implementing this system on the Ethereum testnet, we evaluate its performance in terms of security, scalability, and transaction throughput. Our findings suggest that while the blockchain's inherent features provide robust security, future work is necessary to address transaction efficiency and scalability for larger-scale elections. Optimizing smart contract execution and exploring alternative consensus algorithms are identified as potential solutions for improving system performance in broader applications, such as national elections.

1.1 Blockchain Technology

Blockchain technology's concept was initially introduced in the 2008 Bitcoin whitepaper by Satoshi Nakamoto, 2008 [1], however, curiously, he never used the term "blockchain" in his paper. A mechanism for trustless electronic transactions is suggested in the bitcoin whitepaper. It began with the customary architecture of digitally signed coins, which offers strong ownership control but falls short without a method of preventing double-spending. A distributed network employing proof-of-work for storing the transaction history is suggested as a solution to this problem and if the authentic nodes consume the bulk of CPU power, soon it will make the network computationally prohibitive for an attacker to make changes.

Nodes operate simultaneously and with little cooperation, they don't even need to be recognized as the communications aren't routed to a particular location they are just delivered with the best effort. Nodes can join and quit the network whenever they want, the nodes in a blockchain are blocks, and the link is a cryptographic hash, making it a glorified Linked List. Each block includes a hash for the most recent block, its data, a nonce value, and the previous block's hash. The very first block in a blockchain is named as genesis block where the blockchain is set up. By using a pseudonym as their address, any user can conduct all their transactions in confidence while remaining anonymous to other users, guaranteeing their privacy. Scalability is another feature, allowing any user to sign up whenever they want and begin transactions after downloading the most recent version of the data. Each user can disconnect and rejoin the Blockchain. All they need to do to proceed with a transaction is download all updates since the data's last known state. It offers complete decentralization, which entails that every node has a different copy of every transaction uploaded to the blockchain; this aids in preserving consensus throughout the blockchain.

The functionalities defined in Smart Contracts are used to manage transactions in a blockchain. These decentralized, blockchain-based smart contracts are computer programs. The preceding block's hash and its data are used by miners to solve a puzzle in a proof-of-work fashion. The consensus algorithm is used to assign miners, ensuring that one miner does not perform all of the mining (Fig. 1).

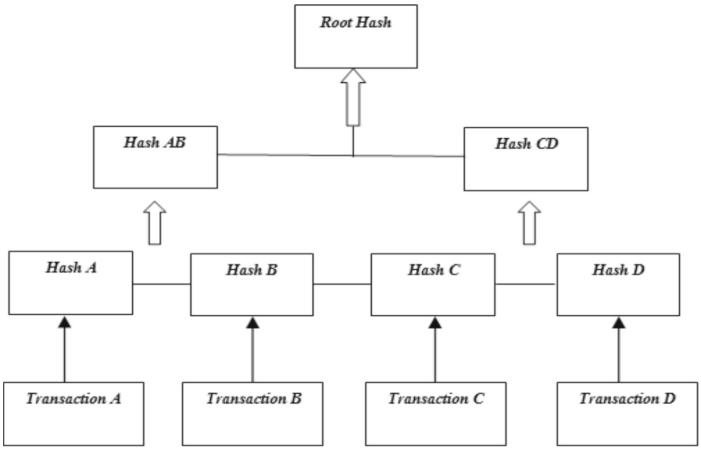


Fig. 1. Merkel Tree

1.2 Motivation

A series of data-containing blocks linked together through a cryptographic hash forms a blockchain. Therefore, any changes that a hacker makes to a single block will affect the entire chain of blocks. This makes the blockchain idea very distinctive.

1. There are several places where the distributed ledger is maintained, eliminating the possibility of a single point of failure.
2. The UN agency will update the ledger with new transactions using distributed administration.
3. To build the changeless chain that gives the blockchain its name, any proposed "new block" to the ledger should refer to the previous version of the ledger without compromising the accuracy of earlier entries.
4. Only after receiving the majority vote from all network nodes can a newly proposed block of entries become a regular component of the ledger.

Following are some distinctive contributions made by the system:

1. Researching blockchain frameworks that can already be used to build smart contracts and electronic voting systems is the first step.
2. The next stage is to propose an electronic voting system that uses a "permissioned blockchain" to modify liquid democracy and is based on blockchain technology.

1.3 Objectives

The objective of this study is to design a voting procedure that meets the following critical criteria:

- Ensure transparency and auditability throughout the voting process.
- Guarantee that each voter's ballot is accurately recorded.
- Restrict ballot casting to eligible electors only.
- Prevent any possibility of tampering or rigging the voting process.
- Safeguard elections from manipulation by any vested interest seeking power.

These criteria will be satisfied by implementing a blockchain-based system that provides:

- **Controlled access:** Restricting ballot casting to currently registered voters.
- **Secrecy:** Ensuring that voters' ballot choices remain anonymous within the system.
- **Accuracy:** Irreversible recording of votes upon casting, ensuring they cannot be altered.
- **Verifiability:** Enabling verification that each cast vote is accurately accounted for by the system.

2 Related Work

Several studies have explored blockchain-based electronic voting systems as a solution to the vulnerabilities found in traditional EVMs. Early systems, such as the one proposed by Ahmed Ben Ayed (2017), utilize blockchain to secure votes by verifying user credentials and storing votes in an immutable chain [3]. However, this system is limited by its lack of flexibility. For instance, if a user mistakenly casts a vote, there is no way to correct the error, leading to potentially inaccurate election outcomes. Additionally, although blockchain prevents vote tampering, this system relies heavily on the security of users' credentials, leaving room for phishing attacks or identity theft.

The "BroncoVote" system proposed by Dagher et al. (2018) leverages Ethereum smart contracts to manage election processes but introduces a server-side cryptographic handling problem [4]. While the use of smart contracts ensures transparency and automation, the need for off-chain encryption to handle large data values introduces a central point of vulnerability. This reliance on external cryptographic operations negates some of the inherent security benefits of decentralization, making the system vulnerable to attacks on the server handling cryptographic functions. McCorry et al.'s (2017) Open Vote Network (OVN) presents another Ethereum-based approach but suffers from limitations in scalability and resistance to malicious miners [11]. The system was only tested with up to 60 voters, and its consensus mechanism is vulnerable to denial-of-service (DoS) attacks.

Further limitations in blockchain e-voting systems have been identified in works like the decentralized anonymous voting system proposed by Lai et al. (2018). While this system enhances voter anonymity and transparency, it lacks a comprehensive auditing mechanism, leaving it vulnerable to external attacks, such as Distributed Denial of Service (DDoS) attacks. Additionally, it struggles to ensure low latency and high throughput, both critical in large-scale elections, where transaction speeds must remain high

to avoid delays in processing results. Finally, Shahzad and Crowcroft's (2019) work on using adjusted blockchain technology to ensure trustworthy voting introduced the BSJC protocol (proof of completeness). While this approach enhances election safety and attempts to address confidentiality and anonymity, the system still suffers from scalability and computational cost issues.

Critical Analysis of Limitations

Across these existing blockchain-based e-voting systems, several limitations remain unaddressed. One common issue is **scalability**. As the number of voters increases, many systems face difficulties in maintaining performance, with delays in vote processing and transaction finality. Solutions like Ethereum struggle with transaction congestion and high gas fees, making large-scale deployment costly and inefficient. Optimizing smart contracts for better transaction throughput remains a key challenge.

Another critical limitation is the **over-reliance on external systems or servers** to handle cryptographic functions, as seen in the BroncoVote system [4]. Off-chain cryptographic processing introduces vulnerabilities that undermine the decentralization and security that blockchain aims to provide. To ensure end-to-end security, future systems must address the integration of fully decentralized cryptographic mechanisms without relying on external servers.

Verifiability is also a recurring issue. Many systems provide individual verifiability but fail to implement robust universal verifiability. Ensuring that voters can verify their vote while maintaining election-wide transparency without revealing individual voting choices remains a technical challenge. Further research is required to develop secure mechanisms that enable both forms of verifiability while safeguarding voter anonymity.

Finally, the **voter experience** is often neglected in these systems. In cases of user error or interface complexity, many systems cannot correct mistakes without compromising security. Implementing user-friendly interfaces and allowing voters to amend their choices before finalizing their votes would significantly enhance usability without introducing new security vulnerabilities.

3 Literature Survey

3.1 Survey Existing System

Adida's study on "Helios" (2008) proposes a rationale for a robust security model and assessment standards aimed at enhancing understanding. Furthermore, it delineates the concept of "Democracy Web Voting," which is more comprehensible compared to the complex "Smart Democracy," the only other theme currently aligning with both the requisite security model and the intended framework [8].

Chaum et al. (2008) introduced integrity, an initial independent end-to-end validation approach that ensures the integrity of optical scan-based voting systems without hindering manual recounts [7]. To promote fairness, Dalia et al. (2012) propose integrating a commitment round. They also suggest a recovery round to facilitate the announcement of election results in cases where voter participation is incomplete. Additionally, it provided an example of ballot secrecy through computational security [9].

The article by Bell et al. (2013) explores the STAR-Voting designs, proposed as a potential next-generation electoral system for Travis County and potentially other jurisdictions [10]. Authors McCorry et al. (2017) introduced the open vote network, a pioneering digital election system utilizing Ethereum, noted for its self-reporting, transparency, and self-tallying capabilities [11]. The structure limited the OVN voting size to only 60 electors. The systematic corruption caused by dishonest miners cannot be mitigated by the OVN. Similarly, dishonest voters can evade the voting process by submitting invalid ballots. The election administrator desires trust, yet the protocol lacks provisions to ensure resilience against coercion or violence (Zhang, S et al. 2020) [12].

Additionally, they needed an additional library to complete the task because Solidity somehow doesn't allow elliptic curve cryptography (Woda, M.; Huzaini, Z., 2019). Once the library was implemented, these generated contracts of voting got too big for storage in the blockchain [14]. According to Hjálmarsson, et al. (2018), OVN is vulnerable to service attacks because of its history of attacks on the Bitcoin network [15]. Lai, W.J.; Hsieh, Y.C.; Hsueh, C.W.; Wu, J.L. 2018 proposal. For DATE—an "A transparent, decentralized e-voting system with low participation confidence requirements." According to researchers, the DATE voting technology can hold significant electronic elections [16].

Shahzad et al. 2019 recommended the BSJC (proof of completeness) as a reliable electronic voting process. They used a process model to explain the overall system's structure, it also attempted, on a smaller scale, to address confidentiality, anonymity, & election safety concerns [17]. But many other difficulties have been raised as well. One such large and challenging mathematical job that requires a great deal of work to accomplish is the proof of labor, when a third party is engaged, there's an additional danger of data manipulation, leaks, and unfair results that could jeopardize end-to-end verification. According to Gao et al. (2019), the polling process may be extended generally by the block's creation and sealing [18].

In a paper, authors Fernández et al., 2020 have presented an audit function-equipped anti-quantum electronic voting method based on blockchain. The code-based approach undergoes modifications to increase its resilience to quantum attacks [19]. The Key Generation Center is a certification authority for certificate-less cryptography (KGC). Besides preserving voter anonymity, it notably simplifies the auditing process. Even in cases where participation in elections is low, a more thorough analysis of their methodology demonstrates that significant security and efficiency benefits remain for an election of this size. To improve security, some efficiency may be decreased if the number is high (Yi, H., 2019) [20].

Torra (2019) proposed enhancing the security of e-voting systems within a P2P network by describing a blockchain-based e-voting scheme [21]. Using a Blockchain Election System based on distributed ledgers could potentially prevent voter fraud. A P2P network was used to design and test the system on Linux computers. Countermeasure-based attacks are a major issue with this method. This approach requires the participation of trustworthy third parties and may not be suitable for centralized implementation in a multi-agent system. Alternatively, employing a distributed approach using secure multiparty computation could offer a resolution. However, the computational

costs could become prohibitive if the computation function is complex and involves numerous participants (K.M. Khan, 2020; Xiao et al., 2019) [22, 23].

3.2 Research Gap

One of the most recent and important technical difficulties facing e-voting systems is secure digital identity management. Before the elections, everybody who wishes to become a citizen needs to sign up to vote. Their information ought to be in a digitally processable format. In addition, any information that involves them should keep their identity information private. The following issues with the outdated E-voting system:

- **Anonymous Voting:** Voters—as well as the system administrators—should remain anonymous after casting their ballots through the system, which may or may not give each candidate an option.
- **Customized voting procedures:** The representation of votes in the pertinent databases or mobile applications is still up for discussion. This is the worst probable approach: a transparent message in text form is more likely to give transparency and integrity than a hashed token. The vote should be regarded as illegitimate in the interim since a symbolic resolution will not be sufficient to obtain it.
- **Electorate-verifiable ballot casting:** Voters need to be able to verify and confirm their votes now they are cast. This ensures awareness and helps prevent or detect any potential malicious activities. Implementing non-repudiation measures can greatly enhance voter confidence. Some contemporary applications partially address these issues autonomously. The Estonian Republic of Estonia should receive a different grade than the others since they have a comprehensive electronic voting system that is equivalent to conventional paper-based elections.
- **Expensive initial deployments, especially for businesses:** Although the operational costs of e-voting systems are considerably lower than those of traditional elections, the initial implementation can incur significant expenses.
- **Growing security issues:** Public opinion polls are seriously threatened by cyberattacks. When an election falls victim to malicious hacking, determining accountability becomes a contentious issue.

4 Proposed Architecture

We created a feasible solution to the problems with the existing e-voting system. The EVM used in the present electronic voting system is quite vulnerable to assault. Our suggested method intends to create a system where every user can vote from any location without fear of interference because the blockchain offers immutability. All the booths that fall under each constituency have access to the secret ledger that is kept. The booths mine the blocks since they are nodes. Every voter uses this system.

- The object-oriented programming language Solidity is used to create smart contracts. On a range of different blockchain platforms, it is utilized to build smart contracts. Since Solidity is a "turning complete" language, loops and conditioned execution statements can be programmed into contracts.

- A full Ethereum node is run and managed via the Geth command line interface. Using Geth has the advantage of allowing the developer to thoroughly test the application before releasing it to the blockchain's main net.
- Ganache is a blockchain emulator for Ethereum. On a machine, it simulates a full Ethereum node. It offers 10 fictitious accounts, each with a balance of 100 ether. Whether or not there are any incoming transactions, the nodes continue mining continuously in the actual world. This heavily utilizes the processing power of the CPU.
- Ganache has the option of auto-mining, where mining only happens when a transaction is received.
- A Dapp browser is required to operate a decentralized application (Dapp). We will use MetaMask to add the capability to the Chrome browser rather than a full-fledged Dapp browser.
- MetaMask was created to make the Ethereum blockchain more user-friendly for the general public. It operates as a Chrome extension, functioning as an Ethereum browser that allows users to interact with decentralized applications and smart contracts without needing to run a full node. This extension also helps manage Ethereum wallets efficiently.
- An application for creating, deploying, and testing smart contracts named Truffle.
- NPM can be utilized to set up the tools and libraries, and Nodejs will be used as a server-side JavaScript platform.
- Before deploying any of the smart contracts into the blockchain, they are all tested using the Mocha test framework.

4.1 System Architecture

The architecture of the proposed system is illustrated in Fig. 2. Users can access the e-Voting App either through a command line interface or a web browser, such as Chrome with the MetaMask plugin embedded.

If accessed via the command line interface, the Node.js console will be displayed to the user, where they need to enter commands directly. Alternatively, users can send commands through a user interface built with HTML, CSS, and JavaScript. These commands are transmitted using Web3.js via RPC(s) and are recognized by Ganache's RPC Endpoint before being executed by the Ethereum VM. The e-voting smart contract developed for this system will be deployed on Ethereum. Various configuration options for the blockchain are managed by the config file.

4.2 System Module

System Setup: Launch a Blockchain as soon as possible. There are two methods to do it:

- Geth: With Geth, we can design and customize our very own Blockchain. Geth is a console that runs on a terminal or Command Prompt and uses Geth tool APIs including web3, personal, admin, and eth to communicate with the blockchain. To obtain a console for the interaction, it leverages the node package management, npm. The user can alter the Blockchain's system status by using these APIs. By transferring

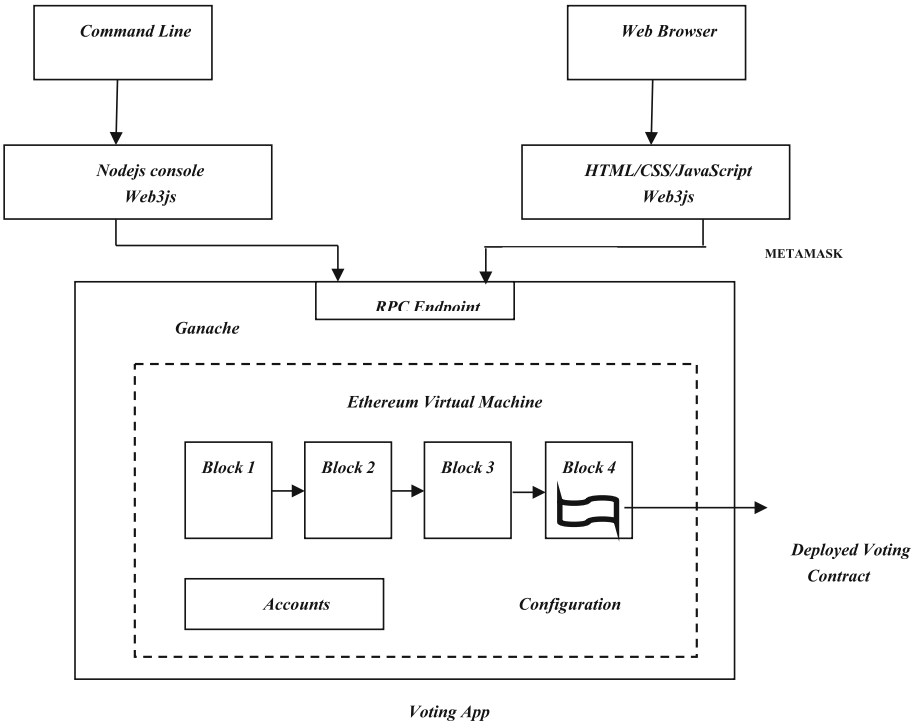


Fig. 2. System Architecture

funds, calling upon contracts, seeing the Blockchain's status, and other features, it can alter the state of the Account. A gateway to the blockchain is provided by this.

- Ganache is a private blockchain tool that provides us with a simulated and visualization environment as well as ten accounts each holding 100 ethers. It uses an RPC network with a specified port and network identifier. Ganache mines automatically; it does not begin mining until an updated transaction has been uploaded to the network. Significant computational resources are saved. It keeps track of every transaction that has taken place as well as every block that has been added to the Blockchain. It has a UI for every interaction, making it simple to use. The configuration of truffle.js is made to connect the Blockchain and the Decentralized application depending on how the Blockchain is running. We must test the Smart Contracts programmed into the system using the Mocha Framework. It statements, assert statements, and JavaScript promises are all used by the Mocha Framework.
- The "it" statement defines the scenario.
- The assert statement for every scenario verifies each condition.
- For any exception that can be foreseen, JavaScript Promises to provide a try-catch framework. All of the smart contracts must be transferred to the deployed Blockchain. Whenever a smart contract migration occurs, the truffle framework is used for migration.

4.3 Voting Process

The process begins with a voter logging in. To do this, he can either create a new account from scratch or import one from Ganache. For each transaction, 100 ethers are granted to the account if it is a newly formed account. The accounts' details are shown on the UI if it is imported. Additionally, that will show how many "ethers" he has in his account, enabling him to check whether he has the necessary number. HTML, CSS, and JavaScript are used to create the user interface. jQuery retrieves the information from the blockchain and sets it to the front end. The list of candidates from which the Voter can choose is seen on the front end. The Name of the Candidate and the Party are shown on the Candidate List. We also have the vote total on hand for debugging purposes. Voting is permitted by the voter. He adds his digital signature to the transaction after casting his vote. The transaction is then announced to the network and mined by the Booths after being signed.

4.4 Outcome Analysis and Display

The results of the elections are tallied by going through each instance of the candidates in the Smart Contract. Using appropriate infographic approaches, the Results are compiled and shown. For this, we import the top libraries, and the JavaScript code is built for this. The outcome is a pie chart that displays the results as a percentage of votes received by each candidate. A preview of the Ganache tool with the private key, setting of GANACHE RPC in remix IDE, a snapshot of the voting button, A confirmation window of the voter casting his vote created by the MetaMask plugin, screenshot of the election result is depicted from Fig. 3. 4, 5, 6, 7 and 8.

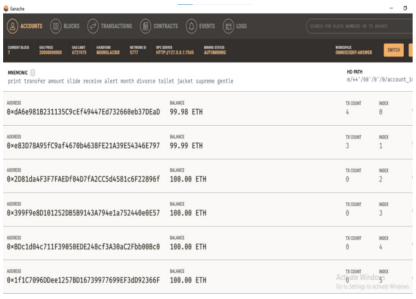


Fig. 3. A preview of the GANACHE Tool running on the network

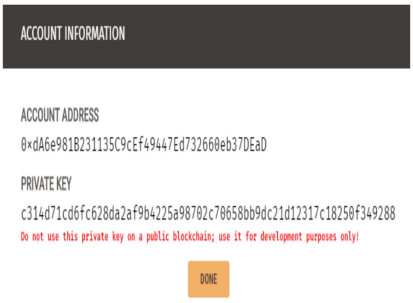


Fig. 4. A preview of the Private key

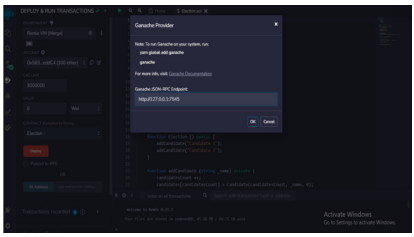


Fig. 5. A preview of setting GANACHE RPC in remix IDE

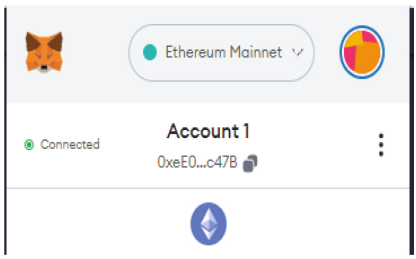


Fig. 6. A confirmation window for a voter to cast his vote is created by MetaMask plugin.

Fig. 7. The snapshot of the voting button.

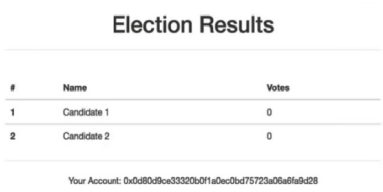


Fig. 8. Screenshot of Election result

5 Conclusion and Future Scope

This study presents a proof-of-concept electronic voting system leveraging Ethereum blockchain technology. Unlike traditional systems, our decentralized approach eliminates the need for trust by enabling eligible voters to securely cast ballots using any internet-connected device. The blockchain's distributed structure ensures resistance to corruption and allows for public verification, thus ensuring transparency throughout the voting process. Voters can verify that their ballots are securely stored on the blockchain while maintaining pseudo-anonymity. Our system, tested on the Ethereum test net, uses smart contracts to manage users, ballots, and related data, enforcing access restrictions effectively. Future research will aim to improve system security and transaction processing efficiency, particularly by exploring consensus algorithms suited to high-security requirements. Additionally, efforts will focus on scaling the system for broader applications, positioning it as a reliable voting solution for universities and similar institutions.

References

1. Nakamoto, S., Bitcoin: a peer-to-peer electronic cash system (2008)
2. Zagórski, F., Carback, R.T., Chaum, D., Clark, J., Essex, A., Vora, P.L.: Remotegrity: design and use of an end-to-end verifiable remote voting system. In: Jacobson, M., Locasto, M., Mohassel, P., Safavi-Naini, R. (eds.) Applied Cryptography and Network Security. ACNS 2013. Lecture Notes in Computer Science, vol. 7954. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-38980-1_28

3. Ayed, A.B.: A conceptual secure blockchain-based electronic voting system. Int. J. Network Secur. Appl. (IJNSA) **9**(3) (2017)
4. Dagher, G.G., Marella, P.B., Milojkovic, M., Mohler, J.: BroncoVote: secure voting system using Ethereum's blockchain. In: Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018), pp. 96–107 (2018)
5. Hardwick, F.S., Gioulis, A., Akram, R.N., Markantonakis, K.: E-Voting with Blockchain: An E-Voting Protocol with decentralisation and voter privacy. ISG-SCC (2018)
6. Kshetri, N., Voas, J.: Blockchain-enabled e-voting. IEEE Software **35**(4) (2018)
7. Chaum, D., et al.: Scantegrity: End-to-end voter-verifiable optical scan voting. IEEE Secur. Priv. **6**(3), 40–46 (2008)
8. Adida, B., 'Helios: Web-based open-audit voting. In: Proceedings of the 17th Conference on Security Symposium, ser. SS'08, Berkeley, CA, USA, pp. 335--348. USENIX Association (2008)
9. Dalia, K., Ben, R., Peter, Y.A., Feng, H.: A fair and robust voting system. In: By Broadcast, 5th International Conference on E-voting, 2012 (2012)
10. Bell, S., et al.: Star-vote: a secure, transparent, auditable, and reliable voting system. In: 2013 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE 13), Washington, D.C. USENIX Association (2013)
11. McCorry, P., Shahandashti, S.F., Hao, F.: A smart contract for boardroom voting with maximum voter privacy. In: Proceedings of the International Conference on Financial Cryptography and Data Security, Sliema, Malta, 3–7 April 2017 (2017)
12. Zhang, S., Wang, L., Xiong, H.: Chaintegrity: blockchain-enabled large-scale e-voting system with robustness and universal verifiability. Int. J. Inf. Secur. **19**, 323–341 (2019)
13. Chaieb, M., Koscina, M., Yousfi, S., Lafourcade, P., Robbana, R.: DABSTERS: distributed authorities using blind signature to effect robust security in e-voting
14. Woda, M., Huzaini, Z.: A proposal to use elliptical curves to secure the block in e-voting system based on blockchain mechanism. In: Proceedings of the International Conference on Dependability and Complex Systems, Wrocław, Poland, 28 June–2 July 2021 (2021). [Google Scholar]
15. Hjálmarsson, F.Þ., Hreiðarsson, G.K., Hamdaqa, M., Hjálmtýsson, G.: Blockchain-based e-voting system. In: Proceedings of the 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), San Francisco, CA, USA, 2–7 July 2018 (2018). [Google Scholar]
16. Lai, W.J., Hsieh, Y.C., Hsueh, C.W., Wu, J.L.: Date: A decentralized, anonymous, and transparent e-voting system. In Proceedings of the 2018 1st IEEE International Conference on Hot Information-Centric Networking (HotICN), Shenzhen, China, 15–17 August 2018 (2018)
17. Shahzad, B., Crowcroft, J.: Trustworthy electronic voting using adjusted blockchain technology. IEEE Access **7**, 24477–24488 (2019)
18. Gao, S., Zheng, D., Guo, R., Jing, C., Hu, C.: An anti-quantum e-voting protocol in blockchain with audit function. IEEE Access (2019)
19. Fernández-Caramés, T.M., Fraga-Lamas, P.: Towards post-quantum blockchain: a review on blockchain cryptography resistant to quantum computing attacks. IEEE Access **8**, 21091–21116 (2020)
20. Yi, H.: Securing e-voting based on blockchain in P2P network. EURASIP J. Wirel. Commun. Netw. **2019**, 137 (2019). [Google Scholar] [CrossRef][Green Version]
21. Torra, V.: Random dictatorship for privacy-preserving social choice. Int. J. Inf. Secur. **19**, 537–543 (2019)
22. Khan, K.M., Arshad, J., Khan, M.M.: Investigating performance constraints for blockchain-based secure e-voting system. Future Gener. Comput. Syst. **105**, 13–26 (2020)
23. Xiao, S., Wang, X.A., Wang, W., Wang, H.: Survey on blockchain-based electronic voting. In: Proceedings of the International Conference on Intelligent Networking and Collaborative Systems, Oita, Japan, 5–7 September 2019

Computer Application for Sustainability

Consumer Usage Intention Factors for Mobile Food- Delivery Applications Using Task Technology Fit Approach

Shweta Shirolkar[1,2(✉)] and Kanchan Patil[3]

[1] Symbiosis International (Deemed University), Pune, India
shwetabagh@gmail.com
[2] Dr Vishwanath Karad MIT World Peace University, Pune, India
[3] Symbiosis Centre for Information Technology, Symbiosis International (Deemed University), Pune, India

Abstract. Consumer decision-making for continuous usage of food delivery apps amid cut-throat competition requires rigorous, disciplined marketing strategies and evaluation. The catalyzing restaurants use online food ordering apps as an alternative channel to reach consumers for delivering the order. This research seeks to verify the dimensions of consumer usage that impact consumer satisfaction and their inclination to continue using online food delivery aggregator applications. Survey questionnaire data from 531 users were collected and spread across the age groups for the analysis and model testing. The results showed a significant impact of Application functionality, Task characteristics, and Task technology fit on consumer satisfaction and continuance intention. The results showed an insignificant association between "perceived ease of use", "perceived usefulness", and "consumer satisfaction". Task Technology Fit significantly affects perceived ease of use, perceived usefulness, and consumer satisfaction. This research study will help practitioners to accord with the competition in the food delivery mobile application sector. The current research also discussed the implications of the findings theoretically and practically with future research agendas.

Keywords: Online food delivery aggregators' application · Task Technology Fit · Technology Acceptance Model · Continuance Intention · Consumer Satisfaction

1 Introduction

Technology has impacted consumer lifestyles and patterns of consumption for a range of goods and services. India ranks 87 out of 120 nations for internet usage, according to the [55] internet usage index. The paradigm shift from e-commerce to m-commerce in online retail has been driven by smartphone technology and internet adoption. In 1997, at the Global Mobile Commerce Forum's first event in Helsinki, Kevin Duffey used the phrase "m-commerce." Later in 1998, it became possible to sell digital content, and there hasn't been a turning back since 2000, when m-commerce became pervasive in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 247–263, 2025.
https://doi.org/10.1007/978-3-031-86305-9_17

digital payments, reservations, voting, etc. In India, smartphone adoption is anticipated to more than double since 2016 to reach 96% by 2040, according to a [56] analysis. [57] Reported 27 billion mobile app downloads in India and other countries. These statistics witness the penetration of smartphones and mobile apps in everyone's life.

[23] define a mobile app as a "software program for small wireless devices like phones and tablets rather than laptops and desktops." Mobile apps are used to shop for almost everything right from apparel, shopping, online ticketing, and food ordering. Smartphone restaurant apps and Online Food Delivery aggregator apps like Zomato, Swiggy, Food Panda, Grub Hub, and others allow clients to order a meal on the go and deliver it to their doorstep. The estimated growth by 2027 is 223.7 billion US dollars, with a CAGR of 11.44 percent.

There are two parts to online food delivery: restaurants serve food to patrons, and food delivery aggregator apps link patrons with restaurants. According to [58], the meal delivery application market is predicted to generate $11.57 mn in sales by 2023, indicating both its growth and opportunity potential. Therefore, it's necessary to apply different concepts and dimensions to analyze how the customer sector consumes and uses online food delivery aggregator services.

1.1 Research Gap

Extant literature on Online Food Delivery aggregator applications focuses on customer experiences, intentions, usage, attitudes, perceptions, barriers, and enablers. The technology acceptance model(TAM) [26] "expectation confirmation model (ECM)", "(UTUAT, UTAUT2)" [6, 36], "Elaboration likelihood model" [34], the service quality model extended TAM [4], Theory of Planned Behavior [14, 41, 45], Value Satisfaction Loyalty Framework [49]; Theory of Consumption Value [22], Extension of Diffusion of Innovation [47], Stimulus- Organism – Response Theory [27, 38],), Innovation Resistance Theory [24], Value Belief Norm theory [21], Theory of Consumer motives [47] are few of the theoretical lenses used in previous research to investigate consumer behavior of online food delivery aggregators applications.

The consumer's continuance of an online food delivery (OFD) aggregator's application is insufficiently analyzed through the theoretical lens of Task Technology Fit (TTF) and its extension. The author believes that Task (food delivery) and technology (OFD aggregator's app functionality) and the compatibility between their characteristics play a crucial role in efficient food delivery tasks that result in consumer satisfaction and loyalty. This research uses TTF with TAM to analyze the new dimensions of usage. The sample for the current study is 531. The sample size of the current study is greater than the previous empirical investigations [8], – 256 respondents, [30], – 367 respondents, that are published on an OFD aggregator's application usage and Continuance intention. The larger sample thus contributes to the broader generalization of the study findings. In their study, [23] asserted that very less no of articles have assessed the mobile applications and their impact on the users of the application. Therefore cur- rent study fills this gap by examining Application functionality. Consequently, the objectives of the current investigation are

1. To examine the dimensions or factors impacting the continuance usage of OFD aggregator applications amongst Indian consumers.

2. To investigate the role of "Task Technology Fit", "Perceived Usefulness (PU)", "Perceived Ease of Use (PEOU)" and "Application Functionality (AF)" on Consumer Satisfaction (SAT) and Continuance Intention (CI) of OFD aggregators applications.

The study determines the aspects or variables that affect India's use of OFD aggregator applications. The study also helps to investigate how AF and TTF affect SAT and CI. By conceptualizing AF as an important construct that influences SAT and CI and makes OFD apps more appropriate for the task of ordering meals.

2 Related Work

2.1 Online Food Delivery Aggregators Application

Mobile application platforms that help customers, merchants, and drivers connect are known as OFD Aggregators [43]. [31] these services are best described as business platforms that do not handle food preparation but do provide order services, payment processing, and process monitoring. The OFD aggregator's operational procedures occur where drivers, restaurants, and customers converge are ensured. OFD aggregators are responsible for enabling restaurants to promptly deliver food orders to consumers' doorsteps within the allotted time.

2.2 Technology Acceptance Model (TAM)

[12] TAM is based on the Theory of Reasoned Action (TRA) [3], which describes behavioral intention as one of the best predictors of consumer usage. TAM is based on PU, or how beneficial a technology is, and improves the task. PEOU describes how easy technology is to use. PEOU and PU affect consumer attitudes toward technology, which affects behavioral intention and usage. [26] applied the TAM to OFD apps, and PU and PEOU predicted app usage intention. The TAM was extended by [45] using TPB. The researchers examined PU, PEOU, utilitarian and hedonic incentives, and food safety risk perceptions that impact attitude, perceived control, trust, and subjective norms influence intention.

2.3 Task Technology Fit (TTF)

[15] proposed the TTF model, which is the "degree of fit between the technology and the task to accomplish the individual tasks and use the technology for the same." This study conceptualizes technology as the AF of OFD aggregator services that determine CI to continue using them. [29] examined food delivery app behavioral intention during COVID-19. The study found that TTF significantly affects app usage behavior. [53] conducted a study that indicated that perceived TTF, trust, performance expectation, confirmation, and social impact affect OFD use intention. Task Technology Fit as a theoretical underpinning was extended by many authors in the past using a different theoretical framework like the "Expectation Confirmation Model (ECM)" to study internet banking continuance [35]. UTAUT was used to study the behavioral intention of library mobile applications [9].

2.4 Task Characteristics and Task Technology Fit (TTF)

Task characteristics are tasks executed by the application of an OFD aggregator. [44] investigated the impact of TTF on task performance in mobile banking. The association between task and TTF is conceptualized within the context of an OFD aggregator's application based on the findings of Tam and Oliveria's study. Therefore,

H1- There is a significant association between task characteristics and TTF

2.5 Application Functionality (AF) and Task Technology Fit (TTF)

[23] posit that very few studies have investigated mobile application features. The role of application features and their influence on consumer adoption and final conversion is also being under-investigated. The current study fills this gap by referring to these application characteristics as AF which is conceptualized as the application's navigation, information accessibility, and aesthetic appeal. [33] Discovered that the features and designs of apps are influenced by the psychological reactions of consumers, which ultimately affect the users of the apps. According to [11] consumers are reluctant to use apps with subpar visual design and appeal. In the present study, AF contributes to technical characteristics and is conceptualized to have a significant association with TTF. Therefore,

H2- There is a significant association between AF and TTF.

2.6 Perception of Ease of Use (PEOU) and Task Technology Fit (TTF)

[10] assessed the association of TTF and PEOU of cloud enterprise resource planning. TTF affected the PU and PEOU of cloud ERP. [25] Examined the role of TTF and TAM in measuring the intention to implement an online project management service. In the study, the TTF of a project management system has a significant impact on its PU and PEOU. Therefore,

H3- There is a significant association between TTF and PEOU for OFD aggregators' applications.

2.7 Task Technology Fit (TTF) and Perceived Usefulness (PU)

[51] conducted a study extending TAM with TTF in which TTF significantly impacts PU. For a massive open online course, [48] examined the relationship between TTF and PU. [40] conducted a study to examine the impact of intrinsic motivation in mobile-assisted language learning. The two intervening variables PU and TTF positively influence students' behavioral intentions. The relationship between TTF and PU in the context of OFD aggregator applications is conceptualized using the preceding arguments and research studies as support. Therefore,

H4-There is a significant association between TTF and PU for OFD aggregators' applications.

2.8 Perceived Ease of Use (PEOU) and Consumer Satisfaction (SAT)

[12] defines PEOU as "the extent to which a person believes using a particular system would be effortless." [1] examined factors affecting SAT and intent to use chatbotsbased services. The study shows that information and service quality improves client happiness. Perceived enjoyment, PEOU, and PU also predict SAT and CI. [37] Studied user happiness and loyalty in online grocery shopping. As per the study findings, PEOU, PU, product assortment, entertainment value, and economic value influence shopper happiness and loyalty. [7] investigated the intention to use e-wallets. Therefore,

H5: PEOU significantly impacts the SAT of OFD aggregators' applications.

2.9 Perceived Usefulness and Consumer Satisfaction (SAT)

The widely known TAM model of technology acceptance [12] defines PU as "the extent to which a person believes that using a system will improve his or her job performance." A study was conducted by [4] to identify the psychological and technological factors that affect consumer conversion to food ordering. The findings supported the assertion that consumers utilize the food ordering app if they find it beneficial. [35] created a unified model to examine the persistent intent of Internet banking. The conceptual model includes technological characteristics, task characteristics, and their influence on SAT and CI. The study also hypothesizes that PU has a substantial effect on SAT, which in turn influences the CI using the product. [45] conducted a study to understand the dimensions of continuous use of the OFD app during the Covid 19. Therefore,

H6: PU has a significant impact on the SAT of OFD aggregator's applications.

2.10 Task Technology Fit (TTF) and Consumer Satisfaction (SAT)

Amidst the novel coronavirus pandemic, [53] investigated the dimensions impacting the continued use of OFD aggregators' applications. Study indicate, TTF has a direct or indirect effect on SAT & continued application usage. [2] conducted a study to determine whether online purchasing would continue during the pandemic. Findings posit that perceived TTF, SAT, PU, and trust affect CI. Therefore,

H7- TTF has a significant impact on SAT in OFD aggregator applications.

2.11 Consumer Satisfaction (SAT) and Consumer Intention (CI)

[28] developed a conceptual framework to comprehend the factors that drive app adoption and continued use among Indian consumers. The model testing confirms that if consumers are pleased with the app, they will continue to use the mobile application.

[46] conducted a study to identify the dimensions of m-service quality that influence user satisfaction and inertia, ultimately resulting in the continued use of the mobile application. The study confirms that user contentment has an impact on the continued use of a mobile application in the given context. Therefore (Fig. 1),

H8: SAT has a significant impact on the CI of OFD aggregator's applications.

Fig. 1. Research Model

3 Research Methodology

Using an online self-administered structured questionnaire primary data is collected. Respondents were selected based on their usage of OFD aggregators' apps in Pune, India – one of the most attractive emerging markets. The selection of the participants is based on purposive sampling. Purposive sampling has been used as a sampling methodology over the random sampling methods as the target respondents decided to be the users of the online food delivery aggregator's apps in the Pune location. The existing literature on online food delivery supports the usage of a non-probability method for selecting the respondents when the target groups are the users of the online food delivery aggregator's application. [39] conducted a study to investigate the "factors affecting online food delivery services in Bangladesh". The study sample was selected using non-probability convenience sampling. [4] surveyed the consumers to know the customer acquisition for food ordering. The study surveyed 374 customers selected based on convenience sampling methods. The data from the survey was analyzed through PLS-SEM. The data collection is done from the total sample size of 531 respondents. The final sample size of the study is derived based on the recommendations of [20]. According to recommendations, the total sample size must be greater than 10 times the constructs linkage in the research model [20]. In the current study, the total inner and out links in the model are 35 which optimizes the sample size to 350 respondents. The data is collected from 531 respondents which justifies the sample size for the current study as it is more than the optimum sample size derived to make the findings generalizable to the larger audience. A pilot study was conducted on 50 cases from the sample and reliability and validity tests were done. Common method bias, as recommended by [19] is conducted for the response bias treatment. Variance inflation factor values of the latent variables and constructs are less than threshold which confirms the absence of common method bias. A 5-point Likert scale is used for the collection of data on the latent variables of all constructs. Figure 2 presents the step-by-step process of methodology adopted for the current study.

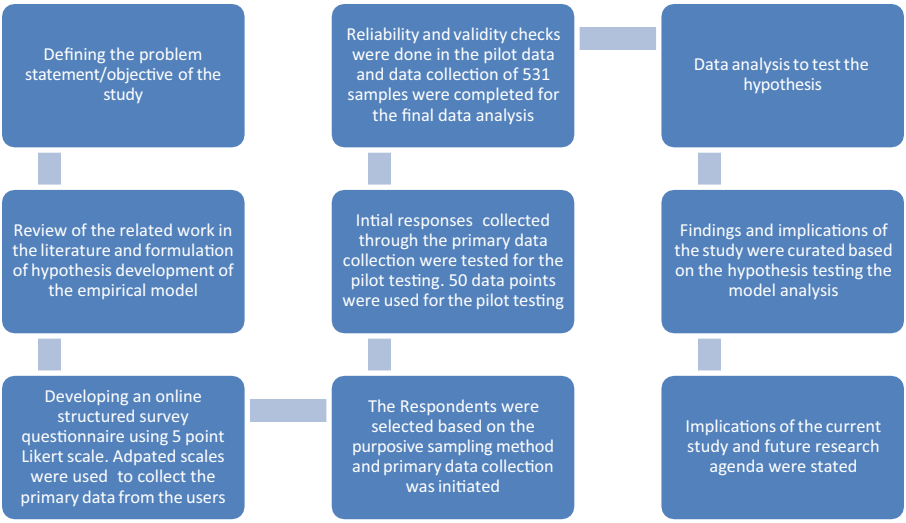


Fig. 2. Flow chart of the Research Methodology

4 Findings

4.1 Respondent Analysis

The sample includes 362 (68%) Males and 169 (32%) Females. Age group composition includes 12 (2%) of the cases under 17 years, 304 (57%) of the cases between 18–22 years, 162 (31%) of the cases between 23–29 years, and 53 (10%) of the cases 30 and above. The frequency of using OFD aggregators apps is distributed as 84 (16%) of respondents with minimal usage (once a month), 94(18%) of respondents with light usage (2–3 times a month), 219 (41%) of respondents with moderate usage (4 times in a month), 71 (13%) with heavy usage (3–4 times in a week), 63 (12%) with extremely often usage (all days of week). Tables 1, 2, and 3 present the tabular format of the demographic profile of the respondents, and Figures 3, 4 and 5 visually present the profile of the

Table 1. Gender Distribution

Gender	Number	(%)
Male	362	68
Female	169	32
Total	531	100

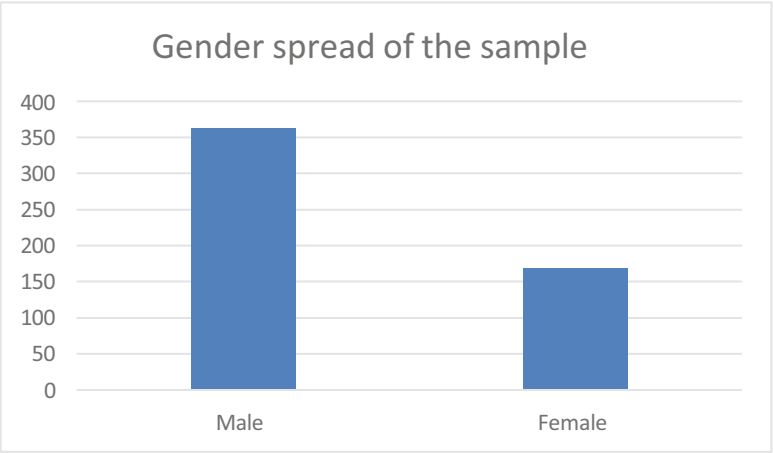


Fig. 3. Gender Distribution

Table 2. Age profile

Age group	Number	(%)
Below 17 years	12	2
18–22 years	304	57
23–29 years	162	31
Above 30 years	53	10
Total	531	100

4.2 Causal Relationships and Model Testing

Measurement Model

All the latent variables of the constructs are significant and have factor loadings of min .70, which is more than .5 [19]. [19] suggested the value of CR to be more than .7 and Cronbach's Alpha near 1 which shows superior internal consistency. The CR and Cronbach's Alpha is at least .7 as presented in Table 4. The AVE value mentioned in Table 4 is more than .5 hence the convergent validity of the constructs is established in the given research model as per the recommendation given by [19]

Discriminant validity is evaluated based on Heterotrait – Monotrait Ratio of Correlations (HTMT). HTMT is the unique approach given by [17]. If HTMT is smaller than one, discriminant validity can be established." Table 5 presents the values of the HTMT for all the constructs.

As presented in Table 6 R^2 is .753 which means 75% of the variance is explained by the input variables on the output variable. Table 7 presents the path coefficients and the significance value of the hypothesis. SAT explained the largest and most significant variance on consumer intention CI (.868) while the variance explained by AF on TTF is lowest (.376) though the relationship between the constructs is significant. The path

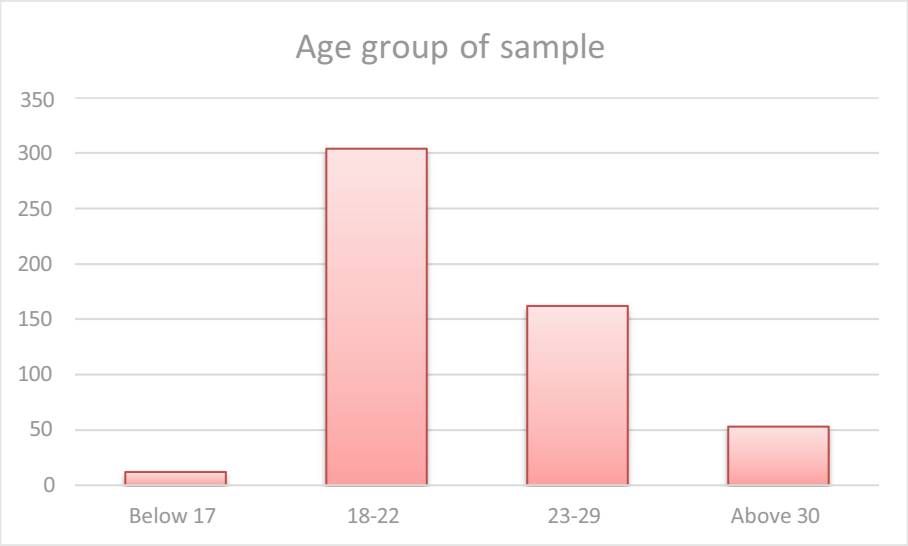


Fig. 4. Age Group

Table 3. Frequency of Usage

Frequency of Usage	Number	(%)
Minimal Usage	84	16
Light Usage	94	18
Moderate Usage	219	41
Heavy Usage	71	13
Extremely Often Usage	63	12
Total	531	100

analysis explains that AF ($\beta = .376$, $p < .01$) and TC ($\beta = .539$, $p < .001$) significantly influence TTF. TTF has a significant positive relationship with PEOU, PU, and SAT ($\beta = .778$, $p < .001$, $\beta = .866$, $p < .001$, $\beta = .697$, $p < .05$) respectively. PEOU ($\beta = -.24$, $p > .05$) and PU ($\beta = .37$, $p > .05$) have an insignificant relationship with SAT. The results of the path analysis support and validate H1, H4, H5, H6, H7, and H8 except for H2 and H3.

5 Implications

The proliferation of smartphones has changed the way consumers purchase goods and services. Whether consumers are purchasing apparel, consumer electronics, transportation, or food services such as food ordering and delivery, the transactions are highly dynamic. The current study investigates CI towards OFD aggregator's apps using the

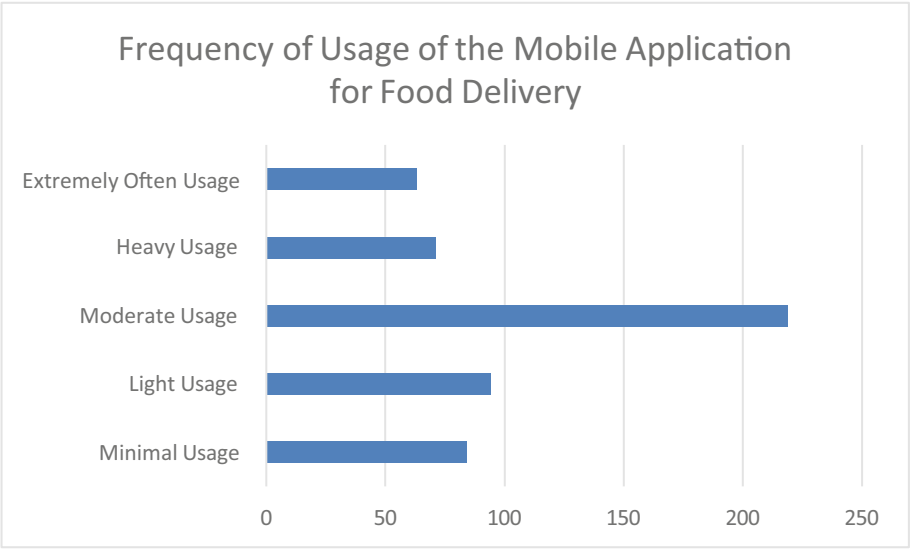


Fig. 5. Frequency of Usage

Table 4. Measurement model results

	Factor Loadings	Average Variance Extracted (AVE)	Composite Reliability (CR)	Cronbach' s Alpha
Application Functionality		.678	.927	.927
AF1	.894			
AF2	.797			
AF3	.799			
AF4	.812			
AF5	.826			
AF6	.81			
Consumer intention		0.746	0.898	0.898
CI1	.892			
CI2	.87			
CI3	.828			
Perception for Ease of Use		0.72	0.885	0.884

(*continued*)

Table 4. (*continued*)

	Factor Loadings	Average Variance Extracted (AVE)	Composite Reliability (CR)	Cronbach' s Alpha
PEOU1	.938			
PEOU2	.791			
PEOU3	.809			
Perception for Usefulness		0.56	0.792	0.79
PU1	.715			
PU2	.771			
PU3	.757			
Consumer Satisfaction		0.67	0.89	0.89
SAT1	.859			
SAT2	.737			
SAT3	.851			
SAT4	.821			
Task Characteristics		0.692	0.871	0.87
TC1	.891			
TC2	.78			
TC3	.821			
Task Technology Fit		0.683	0.915	0.915
TTF1	.852			
TTF2	.813			
TTF3	.795			
TTF4	.798			
TTF5	.871			

theoretical lens of TTF, PEOU, PU, and AF. According to the findings of PLS-SEM, AF, perceived TTF, task characteristics, and SAT explain 75% of the CI [23].

According to Hypothesis H1, task characteristics have a significant relationship with the TTF of OFD aggregators, corroborating with the findings of [44]. This posits that the characteristics of the task of ordering food online like logging into the platform, searching for restaurants, searching for menus, ordering food, and expecting delivery on time, must be complementary with the apps of OFD. AF has a significant relationship with TTF, as supported by previous findings [53, 54]. The relationship suggests that

Table 5. Discriminant Validity- Heterotrait Monotrait Ratio.

	AF	CI	PEOU	PU	SAT	TC	TTF
AF							
CI	0.846						
PEOU	0.699	0.68					
PU	0.794	0.779	0.894				
SAT	0.852	0.868	0.635	0.765			
TC	0.884	0.851	0.71	0.789	0.837		
TTF	0.85	0.888	0.778	0.868	0.834	0.87	

Table 6. R^2

	R^2	R^2 adjusted
CI	.753	.753
SAT	.715	.715
PEOU	.606	.606
PU	.751	.751
TTF	.789	.789

Table 7. Results of Path Analysis

	Path coefficient	Standard deviation (STDE V)	T statistics	P values	Result
AF - > TTF	0.376**	0.131	2.858	0.004	Supported
PEOU - > SAT	0.243ns	0.766	0.318	0.751	Not supported
PU - > SAT	0.377ns	1.007	0.375	0.708	Not supported
SAT - > CI	0.868***	0.027	31.67	0.000	Supported
TC - > TTF	0.539***	0.13	4.146	0.000	Supported
TTF - > PEOU	0.778***	0.033	23.724	0.000	Supported
TTF - > PU	0.866***	0.032	26.866	0.000	Supported
TTF - > SAT	0.697*	0.315	2.21	0.027	Supported

Note: ** p < .01, *** p < .001,* p < .05, ns- not significant

AF – the information quality, visual design, accuracy, relevance, navigation, ordering, etc. should match the task of ordering food online. The relationship between TTF and PEOU (H3) and PU (H4) significantly corroborates the findings of prior studies [10, 40, 53]. TTF has an impact on SAT (H7) and SAT has an impact on the CI (H8) of OFD services, consistent with the findings of previous studies [53] ; [10] means if the task of ordering food online is a good fit with the mobile application it impacts SAT, leading to the CI to use the same platform again. The relationship between PU (H6) and PEOU (H5) with SAT is insignificant, contrary to the researcher's hypothesis and expectation. This contradicts previous research findings [52].

5.1 Theoretical Implications

The current study is one of its kind which extends TTF with TAM to study the CI of OFD aggregator's application. The study conceptualizes Task (food delivery), technology (OFD aggregator's app functionality), and the compatibility between their characteristics plays a crucial role in efficient food delivery tasks that result in SAT and Continuance. It also contributes by theorizing the role of AF as a technology characteristic in TTF, thereby adding a new dimension to TTF in the context of the OFD aggregator's application. AF has a significant impact on TTF, indicating that AF is an important construct to consider while analyzing consumer usage and retention. These findings are similar to the previous studies [53, 54] suggesting that visual design, information quality, service quality, and accuracy are the important factors in ensuring system quality. The research model presents that PEOU and PU are insignificant to impact SAT, which is contrary to the researcher's expectations and previous findings [52]. TTF has a significant impact on SAT, leading to CI. In addition, the TTF significantly impacts the PU and PEOU of the OFD aggregator's application. This is similar to previous studies [10, 40, 53], suggesting the utility and simple use of the mobile app if it fits the task of ordering food online.

5.2 Practical Implications

Practical implications illustrate the factors platform proprietors and app designers of OFD aggregator apps should consider while designing the strategies. SAT with an OFD aggregator's application is dependent on TTF, task characteristics, and AF. TTF has a substantial effect on SAT and, consequently, on CI. Practitioners are required to employ the aforementioned dimensions to monitor SAT and usage factors in the context of OFD aggregator apps to retain customers. Practitioners should persuade consumers that the AF matches the task of ordering food, thereby facilitating SAT. Practitioners are required to strategize and map the different dimensions of AF, such as the ease of downloading the apps, the information on the app, the completeness of information, navigation ease, ease of ordering and payment, incentive options, and order tracking while ordering food online through the mobile applications.

6 Conclusion, Limitations and Future Research Agenda

The study extends TTF by integrating with TAM to investigate an OFD aggregator's application and its continuous usage. In addition to the integrated model the study also conceptualizes the application functions as an important construct impacting the TTF thereafter leading to consumer satisfaction and usage. The study conceptualizes AF impacts TTF and influences SAT and CI. About 75% of the CI is determined by task characteristics, AF, TTF, and SAT. The study also reveals the significant relationship between TTF and PU and PEOU. It contributes theoretically to the literature on OFD aggregator applications, TTF, and TAM. For practitioners, this study suggests that strategists and managers should focus on Task technology fit, application functionality, and task characteristics and match the same with application functionality to boost consumer satisfaction and continuous usage of the application.

The current study has the following limitations which give scope for future research. In future research can enquire the role of TTF and AF for actual usage of apps instead of CI. Multiple population segments from different geographies can be investigated for larger generalizations of the results. The current study estimated a sample size of 531 purposively. Future studies can be conducted by selecting the sample more probabilistically. The current study has a cross-sectional design which can give one-time results for SAT and CI. Longitudinal design can be used to measure the changes in the sample element over some time.

References

1. Ashfaq, M., Yun, J., Yu, S., Loureiro, S.M.C.: I, Chatbot: modeling the determinants of users' satisfaction and continuance intention of AI-powered service agents. Telematics Inform. **54**, 101473 (2020). https://doi.org/10.1016/j.tele.2020.101473
2. Al-Hattami, H.M.: Determinants of intention to continue usage of online shopping under a pandemic: Covid-19. Cogent Bus. Manag. **8**(1) (2021). https://doi.org/10.1080/23311975.2021.1936368
3. Ajzen, I., Fishbein, M.: Understanding Attitudes and Predicting Social Behavior. Prentice-Hall, Englewood Cliffs (1980)
4. Atulkar, S., Singh, A.K.: Role of psychological and technological attributes on customer conversion to use food ordering apps. Int. J. Retail Distrib. Manag. **49**(10), 1430–1446, (2021). https://doi.org/10.1108/ijrdm-09-2020-0349
5. Al Amin, M., Arefin, M.S., Sultana, N., Islam, M.R., Jahan, I., Akhtar, A.: Evaluating the customers' dining attitudes, e-satisfaction, and continuance intention toward mobile food ordering apps (mfoas): Evidence from Bangladesh. European J. Manag. Bus. Econ. **30**(2), 211–229 (2020). https://doi.org/10.1108/ejmbe-04-2020-0066
6. Alalwan, A.A.: Mobile food ordering apps: an empirical study of the factors affecting customer e-satisfaction and continued intention to reuse. Int. J. Inf. Manag. **50**, 28–44 (2020). https://doi.org/10.1016/j.ijinfomgt.2019.04.008
7. Ariffin, S.K., Abd Rahman, M.F.R., Muhammad, A.M., Zhang, Q.: Understanding the consumer's intention to use the e-wallet services. Spanish J. Mark. - ESIC **25**(3), 446–461 (2021). https://doi.org/10.1108/sjme-07-2021-0138
8. Allah Pitchay, A., Ganesan, Y., Zulkifli, N.S., Khaliq, A.: Determinants of customers' intention to use OFD application through the smartphone in Malaysia. Br. Food J. **124**(3), 732–753 (2021). https://doi.org/10.1108/bfj-01-2021-0075

9. Chang, C.-C.: Library mobile applications in University Libraries. Library Hi Tech. **31**(3), 478–492 (2013). https://doi.org/10.1108/lht-03-2013-0024
10. Cheng, Y.-M.: Understanding cloud ERP continuance intention and individual performance: a TTF-driven perspective. Benchmarking: Int. J. **27**(4), 1591–1614 (2020). https://doi.org/10.1108/bij-05-2019-0208
11. Cheung, C.M., Shen, X.L., Lee, Z.W., Chan, T.K.: Promoting sales of online games through customer engagement. Electron. Commer. Res. Appl. **14**(4), 241–250 (2015). https://doi.org/10.1016/j.elerap.2015.03.001
12. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quart. **13**, 319–340 (1989). https://doi.org/10.2307/249008
13. Elvan Dari, C.D., Sukartiko, A.C., Nugrahini, A.D.: Identification of technical requirement for improving quality of local OFD service in Yogyakarta. J. Ind. Inf. Technol. Agric. **1**(2) 1 (2018). https://doi.org/10.24198/jiita.v1i2.14573
14. Gupta, V., Duggal, S.: How the consumer's attitude and behavioral intentions are influenced: A case of OFD applications in India. Int. J. Culture Tourism Hosp. Res. **15**(1), 77–93 (2020). https://doi.org/10.1108/ijcthr-01-2020-0013
15. Good hue, D., Thompson, R.L.: Task-technology fit and individual performance. MIS Quart. **19**, 213–236 (1995). https://doi.org/10.2307/249689
16. Goodhue, D.L., Lewis, W., Thompson, R.: Does PLS have advantages for small sample size or non-normal data? MIS Q. **36**(3), 981–1001 (2012). https://doi.org/10.2307/41703490
17. Henseler, J., Ringle, C., Sarstedt, M.: A new criterion for assessing discriminant validity in variance-based structural equation modelling. J. Acad. Mark. Sci. **43**, 115–135 (2014). https://doi.org/10.1007/s11747-014-0403-8
18. Hair, J.F., Jr., Hult, G.T.M., Ringle, C., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modelling (PLS-SEM). Sage Publications (2016)
19. Hair, J.F., Hult, G.T.M., Ringle, C.M., Sarstedt, M.: A Primer on Partial Least Squares Structural Equation Modelling (PLS-SEM), 2nd edn. Sage, Thousand Oaks (2017)
20. Hair, J.F., Ringle, C.M., Sarstedt, M.: PLS-SEM: indeed a silver bullet. J. Mark. Theory Pract. **19**(2), 139–152 (2011)
21. Hwang, J., Kim, W., Kim, J.J.: Application of the value-belief-norm model to environmentally friendly drone food delivery services. Int. J. Contemp. Hospit. Manag. **32**(5), 1775–1794 (2020). https://doi.org/10.1108/ijchm-08-2019-0710
22. Kaur, P., Dhir, A., Talwar, S., Ghuman, K.: The value proposition of food delivery apps from the perspective of a theory of consumption value. Int. J. Contemp. Hosp. Manag. **33**, 1129–1159 (2021). https://doi.org/10.1108/ijchm-05-2020-0477
23. Kapoor, A.P., Vij, M.: Technology at the dinner table: ordering food online through mobile apps. J. Retailing Cons. Serv. **43**, 342–351 (2018). https://doi.org/10.1016/j.jretconser.2018.04.001
24. Kaur, P., Dhir, A., Ray, A., Bala, P.K., Khalil, A.: Innovation resistance theory perspective on the use of food delivery applications. J. Enterp. Inf. Manag. **34**(6), 1746–1768 (2020). https://doi.org/10.1108/jeim-03-2020-0091
25. Khoa, B.T., Khanh, T.: Online project management service adoption intention: An integration of technology acceptance model and task-technology fit theory. In: 2021 IEEE International Conference on Computing (ICOCO). https://doi.org/10.1109/icoco53166.2021.9673583
26. Lee, W.S., Song, M., Moon, J., Tang, R.: Application of the technology acceptance model to Food Delivery Apps. Br. Food J. **125**(1), 49–64 (2022). https://doi.org/10.1108/bfj-05-2021-0574
27. Liu, Y., Li, Q., Edu, T., Jozsa, L., Negricea, I.C.: Mobile shopping platform characteristics as Consumer Behavior Determinants. Asia Pacific J. Mark. Logist. **32**(7), 1565–1587 (2019). https://doi.org/10.1108/apjml-05-2019-0308

28. Malik, A., Suresh, S., Sharma, S.: Factors influencing consumers' attitude towards adoption and continuous use of mobile applications: a conceptual model. Procedia Comput. Sci. **122**, 106–113 (2017). https://doi.org/10.1016/j.procs.2017.11.348
29. Muangmee, C., Kot, S., Meekaewkunchorn, N., Kassakorn, N., Khalid, B.: Factors determining the behavioral intention of using food delivery apps during COVID-19 pandemics. J. Theoret. Appl. Electron. Comm. Res. **16**(5), 1297–1310 (2021). https://doi.org/10.3390/jtaer16050073
30. Okumus, B., Ali, F., Bilgihan, A., Ozturk, A.B.: Psychological factors influencing customers' acceptance of smartphone diet apps when ordering food at restaurants. Int. J. Hosp. Manag. **72**, 67–77 (2018). https://doi.org/10.1016/j.ijhm.2018.01.001
31. Pigatto, G., Machado, J.G.D.C.F., Negreti, A.D.S., Machado, L.M.: Have you chosen your request? Analysis of online food delivery companies in Brazil. Br. Food J. **119**(3), 639–657 (2017). https://doi.org/10.1108/bfj-05-2016-0207
32. Pandey, S., Chawla, D., Puri, S.: Food delivery apps (FDAS) in Asia: an exploratory study across India and the Philippines. Br. Food J. **124**(3), 657–678 (2021). https://doi.org/10.1108/bfj-01-2020-0074
33. Peters, T., Işık, Y., Tona, O., Popovič, A.: How system quality influences mobile bi use: the mediating role of Engagement. Int. J. Inf. Manag. **36**(5), 773–783 (2016). https://doi.org/10.1016/j.ijinfomgt.2016.05.003
34. Pillai, S.G., Kim, W.G., Haldorai, K., Kim, H.S.: OFD services and consumers' purchase intention: integration of theory of planned behavior, theory of perceived risk, and the elaboration likelihood model. Int. J. Hosp. Manag. **105**, 103275 (2022). https://doi.org/10.1016/j.ijhm.2022.103275
35. Rahi, S., Abd Ghani, M.: Examining Internet banking user's continuance intention through the lens of technology continuance theory and task technology fit model. Digital Pol. Regul. Govern. **23**(5), 456–474 (2021)
36. Ramos, K.: Factors influencing customers' continuance usage intention of food delivery apps during COVID-19 quarantine in Mexico. Br. Food J. **124**(3), 833–852 (2021). https://doi.org/10.1108/bfj-01-2021-0020
37. Sreeram, A., Kesharwani, A., Desai, S.: Factors affecting satisfaction and loyalty in online grocery shopping: an integrated model. J. Indian Bus. Res. **9**(2), 107–132 (2017). https://doi.org/10.1108/jibr-01-2016-0001
38. Shah, A.M., Yan, X., Shah, S.A.A., Ali, M.: Customers' perceived value and dining choice through mobile apps in Indonesia. Asia Pacific J. Mark. Logist. **33**(1), 1–28 (2020). https://doi.org/10.1108/apjml-03-2019-0167
39. Saad, A.T.: Factors affecting online food delivery service in Bangladesh: an empirical study. Br. Food J. **123**(2), 535–550 (2020). https://doi.org/10.1108/bfj-05-2020-0449
40. Sun, Y., Gao, F.: An investigation of the influence of intrinsic motivation on students' intention to use mobile devices in language learning. Educ. Technol. Res. Dev. **68**(3), 1181–1198 (2019). https://doi.org/10.1007/s11423-019-09733-9
41. Troise, C., O'Driscoll, A., Tani, M., Prisco, A.: OFD services and behavioral intention – a test of an integrated TAM and TPB framework. Br. Food J. **123**(2), 664–683 (2020). https://doi.org/10.1108/bfj-05-2020-0418
42. Terhorst, Y., et al.: Validation of the mobile application rating scale (MARS). PLoS ONE **15**, 0241480 (2020). https://doi.org/10.1371/journal.pone.0241480
43. Tong, S., Luo, X., Xu, B.: Personalized mobile marketing strategies. J. Acad. Mark. Sci. **48**, 64–78 (2020)
44. Tam, C., Oliveira, T.: Understanding the impact of M-banking on individual performance: Delone & McLean and TTF perspective. Comput. Hum. Beh. **61**, 233–244 (2016). https://doi.org/10.1016/j.chb.2016.03.016

45. Wen, H., Pookulangara, S., Josiam, B.: A comprehensive examination of consumers' intentions to use food delivery apps. Br. Food J. **124**(5), 1737–1754 (2021)
46. Wang, W.-T., Ou, W.-M., Chen, W.-Y.: The impact of inertia and user satisfaction on the continuance intentions to use mobile communication applications: A Mobile Service Quality Perspective. Int. J. Inf. Manag. **44**, 178–193 (2019). https://doi.org/10.1016/j.ijinfomgt.2018.10.011
47. Wang, O., Scrimgeour, F.: Consumer adoption of online-to-offline food delivery services in China and New Zealand. Br. Food J. **124**(5), 1590–1608 (2021). https://doi.org/10.1108/bfj-03-2021-0208
48. Wu, B., Chen, X.: Continuance intention to use MOOCs: integrating the Technology Acceptance Model (TAM) and task technology fit (TTF) model. Comput. Hum. Behav. **67**, 221–232 (2017). https://doi.org/10.1016/j.chb.2016.10.028
49. Xu, C., Peak, D., Prybutok, V.: A customer value, satisfaction, and loyalty perspective of mobile application recommendations. Dec. Support Syst. **79**, 171–18 (2015). https://doi.org/10.1016/j.dss.2015.08.008
50. Yusra, Y., Agus, A.: OFD service quality: does personal innovativeness matter? Asia Proc. Soc. Sci. **2**(3), 251–255 (2018). https://doi.org/10.31580/apss.v2i3.453
51. Yen, D.C., Wu, C.S., Cheng, F.F., Huang, Y.W.: Determinants of users' intention to adopt wireless technology: An empirical study by integrating TTF with Tam. Comput. Hum. Behav. **26**(5), 906–915 (2010). https://doi.org/10.1016/j.chb.2010.02.005
52. You, J.-J., Jong, D., Wiangin, U.: Consumers' purchase intention of organic food via social media: the Perspectives of Task-technology fit and post-acceptance model. Front. Psychol. 11 (2020). https://doi.org/10.3389/fpsyg.2020.579274
53. Zhao, Y., Bacao, F.: What factors determining customers' continuing using food delivery apps during the 2019 Novel coronavirus pandemic period? Int. J. Hosp. Manag. **91**, 102683 (2020). https://doi.org/10.1016/j.ijhm.2020.102683
54. Zhou, T., Lu, Y., Wang, B.: Integrating TTF and UTAUT to explain mobile banking user adoption. Comput. Hum. Behav. **26**(4), 760–767 (2010). https://doi.org/10.1016/j.chb.2010.01.013
55. Topic: Internet Usage in India. Statista (2023). www.statista.com/topics/2157/internet-usage-in-india. https://www.statista.com/topics/2157/internet-usage-in-india/
56. Smartphone Penetration Rate in India 2009-2040." Statista (2023). www.statista.com/statistics/1229799/india-smartphone-penetrationrate. https://www.statista.com/statistics/1229799/india-smartphone-penetration-rate/
57. App Downloads 2021–2022, by Country. Statista (2023). www.statista.com/statistics/1287159/app-downloadscountry. https://www.statista.com/statistics/1287159/appdownloads-by-country/
58. "Online Food Delivery - Worldwide | Statista Market Forecast. Statista, www.statista.com/outlook/dmo/online-food-delivery/worldwide. https://www.statista.com/outlook/dmo/online-food-delivery/worldwide

Optimizing Query Execution Time in Relational Databases Using Cost-Based and Index Merge Strategies: A Case Study on E-Commerce Data

Patel Nilen[1(✉)] and Himansu Maniar[2]

[1] Bhagwan Mahavir Center for Advance Research, Bhagwan Mahavir University, Surat, Gujarat, India
nilenp22@gmail.com

[2] Bhagwan Mahavir College of Computer Application, Bhagwan Mahavir University, Surat, Gujarat, India

Abstract. A good in-depth analysis of query execution time with optimization in relational databases, focusing on mainly two strategies which are cost-base and index merge by using e-commerce as a case study. In this era, the data is increasing with its complexity in e-commerce which needs efficient query optimization techniques that ensure accurate data retrieval by reducing the time. We explore the main principles of cost-based optimization, which involves the cost of different query execution plans evaluation and selecting the most efficient plan. In addition, we examine the index merge technique that reduces query execution time using multiple indexes concurrently. By doing various experiments on e-commerce databases, we show the improvements in query execution time achieved by these strategies. The result focuses on the best effectiveness of these two strategies in reducing resource usage and enhancing performance of relational database management systems. Our findings help database administrators and developers who are seeking to optimize query performance in large scale database or data-intensive applications.

Keywords: Query Optimization · Cost-Based Strategies · Index Merge Strategies · Query Performance Analysis

1 Introduction

In this era of e-commerce and big data, efficient query processing in relational database management system has become the most critical for business management to ensure user satisfaction and stay more reliable in this competitive world. Data is continuously growing with complexity, mostly in e-commerce environments, there is need for advanced query optimization techniques has become more dominant [9, 13, 17].

Query optimization is a crucial component of the database management systems, which is aimed to improve query performance in order to reduce query execution time and also resource consumption [7, 15, 18]. Approaches that have been followed in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 264–273, 2025.
https://doi.org/10.1007/978-3-031-86305-9_18

traditional query optimization often struggle with complexities of modern workloads of analytics that commonly found in e-commerce databases [16].

Cost-based query optimization has appeared as a strong technique for improving performance by the evaluation of different execution plans and selecting the most efficient plan based on estimated costs [3, 6, 15]. This is the approach that takes into account facts like available indexes, distribution of data and system resource in order to make informed decisions about query execution strategies [1, 3, 8].

Another encouraging strategy to optimize query is the use of index merge techniques, which can noticeably enhance query performance by combining multiple indexes efficiently [8, 17]. This approach is specifically effective in scenarios where multiple conditions need to be satisfied, which is common event in e-commerce databases with various user preferences and complex product attributes.

Regardless of the advancements in techniques of query optimization, their application in the context of the e-commerce data that presents unique challenges because of the dynamic nature of the data, diversity in patterns of query, and the need for real-time performance [10, 12]. In addition, an investigation on the integration of index merge strategies and cost-based optimization needs to be done.

The aim of this research paper is to investigate the effectiveness of combining index merge strategies and cost-based optimization in order to improve query execution time in relational database, with focus on e-commerce data. By operating a comprehensive case study, we pursue to provide observational evidence of the benefits and limitation of these approaches in a real-world e-commerce context.

Our study builds on the existing research on query optimization [2, 4, 5, 11, 14] and extends it by addressing the specific challenges which is put forward by e-commerce data. We will explore how these strategies can be built custom in order to handle the unique characteristics of e-commerce queries, such as complex join operations, aggregations and periodic updates [16, 17].

2 Literature Survey

Over the years it has been observed that the field of query optimization in relational databases has been significant advancements, with numbers of studies that focuses on improving query execution time and efficiency. Cost-based query optimization, initiated by Selinger et al. in their work on System R [3], laid on foundation for modern query optimizers by introducing the concept of evaluation of multiple execution plans of query based on estimated costs. This is the approach which has been further refined and developed by other researcher like Baril and Bellahsène [1], who proposed different methods for selecting materialized views using cost-based approaches. Their work demonstrated remarkable performance improvements using TPC-H benchmark schema which focus on the potential of cost-based optimization in complex database environments.

In the domain of distributed systems, researchers like Adali et al. [5], who have addressed that the challenges in optimization of queries that can be executed in abstraction layer between databases and applications, introducing concepts like intelligent caches and invariants to enhance the use of cached results. This work has relevancy to e-commerce systems that mostly depend on data sources which are distributed. Index

merge techniques have also gained notability as an effective way to optimize performance of query, especially for complex queries which have multiple conditions. Observational studies, such as those conducted by Mehta et al. [13], have point up the benefits of combining multiple indexes in order to improve query execution time.

The unique challenges that are presented by e-commerce database, rendered by their complex patterns of different queries and dynamic nature which have been addressed in different research efforts. Wei Cao et al.'s work [10] on the Unified MySQL Platform (UMP) at Taobao offers most valuable insights into optimizing database performance in large-scale e-commerce environments, addressing issues such as read/write splitting, database sharding and ensuing high availability. In same way, the DeCloud-RealBase framework proposed by Wei Cao et al. [10], while not specifically focused at e-commerce, provides relevant approaches in order to handle real-time data writing and querying, which are very critical in dynamic environments of e-commerce.

Latest advancement in optimization of query have introduced more refined techniques. The work of Paudel and Bhatta [15] provides comprehensive correlation of different algebra queries, featuring the importance of minimizing disk access in centralized relational databases. Niu et al.'s research [14] on integrating the Orca optimizer into MySQL which demonstrate the potential for significant improvements in performance with respect to handling complex analytical queries, which have relevancy for e-commerce system particularly that often require both analytical and transactional query processing.

Research on query optimization in uncertain and probabilistic databases which not directly connected to the e-commerce databases, such as the work by Kheradkar and Shirgave [18], offers valuable analysis which could be applicable to specific aspects of e-commerce data analysis. Their exploration of different models and techniques provides a framework for handling the essential uncertainties in real-world data. On the far side of query-level optimization, studies have also focused on system-level tuning in order to improve database performance, highlighting the importance of tuning SQL queries, optimizing server and instance parameters and considering hardware design in overall system performance [8].

Investigations into techniques and strategies aimed at enhancing the performance and efficiency of queries that trace the origin within a database, such as that conducted by Galvic et al. [14], introduces new techniques for instrumented queries. While their main focus is on tracking data lineage, the proposed framework for query transformation and optimization could potentially be adopted for e-commerce scenarios, mainly in areas which require detailed tracking of data modifications and query histories. In addition, as e-commerce systems often operate in distributed environments, research on the evolution of query optimization methods from centralized to data grid systems [7] provide insights into the challenges and solutions for query optimization in distributed and complex environments which is valuable.

Even with these major developments, the remaining gap in the literature specifically addressing the unique challenges of e-commerce databases, particularly in combining two strategies which are cost-based optimization and index merge strategies. Our research that aims to bridge the gap by building upon these foundational works and

extending their application with respect to the e-commerce data. By integrating cost-based optimization with index merge strategies and by considering the unique characteristics of e-commerce queries, we seek to contribute to the ongoing efforts in order to enhance database performance in this critical sector, specifically opening new avenues for research and practical applications in the fields of query optimization for e-commerce databases.

3 Proposed Methodology

The proposed methodology which focuses on optimizing MySQL query execution time in e-commerce databases which supports a combination of cost-based optimization and index merge strategies. At the heart of this approach is that the strategic use of MySQL's "optimizer_switch" setting, individually focusing on index merge capabilities. The methodology outlines our approach which is given below:

- Database Environment Setup: We begin by setting up e-commerce database environment using MySQL. This database will be settled by different set of tables which mimics real-world e-commerce data structure. This database includes tables like products, customers, orders, payments and reviews. The data distribution and volume will be carefully designed to reflect e-commerce scenarios, with specific attention to relationship complexities.
- Baseline Performance Measurement: Prior to optimization, we will create a baseline performance metrics for a set of representative queries. These queries will be designed to cover a wide range of operations of e-commerce which includes product searches, order details, payment details and analytics. Execution time and resource utilization will be recorded for each query under normal operating conditions.
- Optimizer Switch Configuration: The core of our optimization strategy is around the precise configuration of MySQL's "optimizer_switch". We will use the configuration which is given below:

SET optimizer_switch = 'index_merge = on, index_merge_union = on, index_merge_sort_union = on, index_merge_intersection = on, index_merge_sort_intersection = on';

This configuration enables various index merge strategies:

o index_merge: Allows the optimizer to consider merging multiple indexes.
o index_merge_union: Enables the union of index scans.
o index_merge_sort_union: Permits sorted union of index scans.
o index_merge_intersection: Allows intersection of index scans.
o index_merge_sort_intersection: Enables sorted intersection of index scans.
o Index Design and Implementation: Based on the query patterns identified in step 2, we will design and implement a comprehensive indexing strategy.

- Analysis Cost-based Optimization: For each query, we will conduct a cost-based optimization analysis which includes utilizing MySQL's EXPLAIN ANALZE command in order to gather execution plans and statistics. After that by analysing optimizer's choices in respect of index usage and merge strategies. Then identify in which situation, the optimizer chooses index merge strategies over other options.

- Performance Testing and Measurement: We run extensive performance on the optimized query by executing each query multiple times under different database loads in order to ensure consistent results. While performing these actions, it is necessary to measure and record the execution times, CPU usage, I/O operations and memory consumption. After that compare these metrics against the baseline measurements from step number 2. Identifying queries that show the most remarkable improvements, that may require further optimization.
- Documentation and Reporting: Throughout the process, we will maintain detailed documentation of all configuration changes and impacts, query transformation and their effect on performance, comparative analysis of before and after optimization states.

This comprehensive methodology aims to provide a robust framework for optimizing query execution time in MySQL for e-commerce databases. By systematically applying cost-based optimization techniques and leveraging index merge strategies, we expect to achieve significant performance improvements across a wide range of e-commerce query patterns. The results and insights gained from this study will contribute valuable knowledge to the field of database optimization, particularly in the context of dynamic and complex e-commerce environments.

4 Results

The implementation of our cost-based and index merge optimization strategies in the MySQL database for e-commerce data has promising results. We conducted performance tests on a set of multiple representative queries, comparing their execution times before and after applying optimization techniques.

After applying our optimization techniques, which included configuration of MySQL's optimizer_switch that enable various index merge strategies, along with query rewriting and index design, we observed notable improvements in query execution times (Fig. 1).

Baseline Measurements	60.9983ms	60.2138ms	83.8581ms	69.9596ms	73.0873ms
After Optimization	60.1431ms	59.2999ms	77.8760ms	57.3840ms	66.9344ms

Fig. 1. Comparison of Baseline Measurements and After Optimization

By analysing these results, we can see that these queries showed improvements in execution time. First two results shows that there is reduction in query execution time of about 1.4% and 1.5% respectively. Third result shows 7% decrement in execution time. While forth result shows improvement of approximately 18% which is remarkable. Fifth result shows 8.4% reduction of query execution time.

On average, across all queries, we observed a 7.8% reduction in query execution time specifically in the context of e-commerce database. This small reductions in query

execution time can lead to substantial improvements in overall system performance and user experience.

The outcomes presented in this study confirm the applicability of the proposed approach in reducing the query response time by applying cost-based and index merge optimization strategies. That all types of queries have been improved steadily indicates that the proposed methodology can be effectively used in virtually any e-commerce database.

However, it was observed that the level of improvement was high for some queries but significantly low for others. Thus, these fluctuations on query performance stress how profound an issue query optimization is and how specific methodological approaches are likely to succeed depending on the general shape of the query as well as the nature of data structures.

This approach of optimizations has proven to have concrete improvements to cutting down the query execution time involving e-commerce data. Altogether, these changes were between 1. 4% to 18%, should be marked as a rather considerable improvement in the database performance, and, correspondingly, in the response of specific e-commerce applications.

5 Discussion

From our investigation on minimizing query response time in using cost-based and index merge approaches for e- commerce data in relational databases; one can learn a number of critical lessons and impacts on DBMS in e-commerce.

- Effectiveness of Optimization Strategies: When discussing the cost-based optimization and index merge strategies, one can claim that our implementation yielded constant performance gains for every tested query, with the speedup factor ranging from 1 to 7. This variability of improvement indicates that the benefit of the identified strategies is query-dependent. The results show that in the first query, which involved simple select and join statements, there was not as much of an improvement as we observed in the second and fourth queries, where our optimization approach was most effective for more complex joins and queries with multiple conditions. This is in accordance with Mehta et al. [13] who rightly pointed out that for a number of queries one or more indexes are involved.
- Impact on E-commerce Performance: In e-commerce applications, database manages many transactions and queries at the same time, hence, achieving even few milliseconds of improvement in such event will result into great gain in total system performance. The potential of improving execution time by 8% for the test queries we implemented can imply tremendous enhancements of overall system performance, notably during frequent surge in traffic. This improvement is as per the objective available in other study on the enhancement of large-scale e-commerce database performance.
- Scalability Considerations: Although our results look quite positive, we need to assess how these optimizations behave with growing data size and query intricacy. The inconsistent level of enhancement observed in different queries indicates that the extensibility of our solution may be contingent on the peculiarities of queries and

corresponding data structures. For the future work, the authors should consider the evaluation of these optimizations while considering different loads and large datasets to prove its applicability on growing e-commerce platforms.

- Trade-offs in Optimization: The relatively small jumps in the results for some of the queries (for example 1. 4% and 1. 5% for the first two queries) show another important aspect of query optimization: the law of diminishing returns. Lower queries make additional gains progressively more challenging to obtain. This observation submits that there is need for optimization to consider the objective of optimization in proportion to the cost necessary to achieve optimization.
- Implications for Index Design: It also indicates that the design of the indexes plays a crucial role in determining the performance of the queries on them. But the creation and maintenance of multiple indexes involve storage and update overhead costs to the database system. Subsequent studies should investigate the behavioral effects that come with large indexing and its effects on the all-round performance of the database, especially when used in a write-intensive e-commerce environment.
- Comparison with Existing Approaches: In the following, we further develop the idea presented by Selinger et al. [3] regarding cost-based optimization in the e-commerce data scenario. The changes that we have made indicate that our method can adapt to some of the concerns that are characteristic of e-commerce databases – that is, concurrency control and support for both transactional and analytical processing as noted by Niu et al. [16]
- Limitations and Future Directions: Although, our study outlines the prospects of cost-based and index merge strategies, the study has some limitations. It is also important to note that the test queries we used may not span all the query pattern possibilities in e-commerce systems. The further work should enrich the types of queries for experimental evaluations and investigate the effectiveness of the introduced optimizations in the context of e-commerce-specific scenarios like flash sales or seasonal traffic.

Also, in the present research, we largely concentrated on the read operation of our memory hierarchy. Since e-commerce data is constantly changing, additional research work should be done on analysing the effects of the above optimization strategies on write-intensive operations and the best strategies for ensuring efficient data writes as new data arrives.

Practical Implications: For database administrators and e-commerce platform developers, our findings highlight the importance of leveraging advanced query optimization techniques. The configuration of MySQL's optimizer_switch, as demonstrated in our methodology, provides a practical approach to enhancing query performance. However, the variability in improvement across queries emphasizes the need for careful analysis and customization of optimization strategies for each specific e-commerce environment.

- Theoretical Contributions: Our study contributes to the theoretical understanding of query optimization in relational databases by demonstrating the combined effect of cost-based optimization and index merge strategies in a specific domain (e-commerce). The objective models that we have described in this study including the Index Efficiency Metric (IEM) and the Optimization Effectiveness Score (OES), present new ways of benchmarking optimization approaches in future studies.

6 Conclusion

In this research, cost-based optimization and index merge approaches have been discussed in an endeavour to enhance the use of query response time in relational databases with emphases to e-commerce data. From our study, we have the following findings and contributions to the field of database optimization.

When using our optimization techniques, we observed consistent performance enhancements in all the queries that we experimented on, with the execution time being cut by as much as 1. Approximately 8% of the query execution time was saved. These results show the effectiveness in using the cost-based and index merge strategies in optimizing e-commerce database.

As shown above, the proposed approach involving tuning the optimizer_switch in MySQL and proper index design was efficient for most query types used in most e-commerce systems. Such consistency implies that the proposed methodology can be applied in a variety of different e-commerce database contexts.

The discussed fluctuation of improvement in different queries proves that query optimization in e-commerce environments is a multifaceted process. It highlights the importance of query specific optimization methodologies that take into account the nature of the queries being posed as well as the nature of the underlying data structures.

Despite the fact that the study was conducted using a limited set of queries and a specific e-commerce database configuration, the research and methodologies discussed herein can be applied in future studies and real-world environments. These optimization techniques are further researched and fine-tuned due to the positive outcomes seen from the study.

However, one should remember that query optimization is the process that keeps on going. Thus, it will be seen that future trends in e-Commerce and the increase in data volumes will require changes to optimization strategies as well. The information described in this work can be seen as a foundation for further development in the area of e-commerce database systems and their optimization for more effective functioning; however, an indication is made to the continued need for research and improvement in this particular field.

Therefore, this paper proves that pursuing the ideas of cost-based optimization and index merge, the opportunity to improve the query's speed of e-commerce databases, is provided. Though, the degree of improvement differs, what can be seen is that there is a possibility that application of these techniques may help to enhance the efficiency of data search and analysis in e-commerce systems. With the e-commerce sector becoming a major driver for today's business growth, development and deployment of such optimization strategies are likely to be a critical factor determining the scalability of database systems in delivering the expected speed and efficiency.

7 Future Scope

Query Diversity and Scalability: Expand the study which includes a wider range of query types, complexities which reflects the diverse nature of e-commerce transactions. Investigate the scalability of the proposed optimizations for higher query volumes and larger datasets.

Long-term Index Management: Develop adaptive index management strategy that evolves with changing query patterns and data distributions over time. Explore the impact of these strategies on long-term database performance and maintenance.

Machine Learning Integration: Investigate the potential of machine learning techniques to predict optimal query execution plans and automatically adjust optimization strategies based on historical performance data.

Real-time optimization: Find strategies for real-time query optimization that can adapt to sudden changes in workload or statistical characteristics during high e-business periods.

References

1. Baril, X., Bellahsène, Z.: Selection of Materialized Views: A Cost-Based Approach
2. Wong, E., Youssefi, K.: Decomposition—a strategy for query processing. ACM Trans. Database Syst. **1**(3), 223–241 (1976), https://doi.org/10.1145/320473.320479. Accessed 26 Jul 2024
3. Griffiths, P., et al.: Access Path Selection in a Relational Database Management System
4. Gardarin, G., et al.: Calibrating the Query Optimizer Cost Model of IRO-DB, an Object-Oriented Federated Database System*
5. Adali, S., et al.: Query caching and optimization in distributed mediator systems. In: Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data - SIGMOD '96 (1996). https://doi.org/10.1145/233269.233327. Accessed 26 Jul 2024
6. Zhu, Q., et al.: Cost estimation for queries experiencing multiple contention states in dynamic multidatabase environments. Knowl. Inf. Syst. **5**(1), 26–49 (2003). https://doi.org/10.1007/s10115-002-0070-9. Accessed 26 Jul 2024
7. Hameurlain, A.: Evolution of query optimization methods: from centralized database systems to data grid systems*. In: LNCS, vol. 5690, pp. 460–470 (2009). Accessed 26 Jul 2024
8. Karthik, Mr., et al.: Tuning the SQL Query in Order to Reduce Time Consumption
9. Hamdoon, S., et al.: Pragmatic approach to query optimization general terms. Int. J. Comput. Appl. **66**(7), 975–8887 (2013). Accessed 26 Jul 2024
10. Cao, W., et al.: Realization of the low cost and high performance MySQL cloud database. Proc. VLDB Endowment **7**(13), 1742–1747 (2014). https://doi.org/10.14778/2733004.2733077. Accessed 26 Jul 2024
11. Lupher, A.: Shark: SQL and Analytics with Cost-Based Query Optimization on Coarse-Grained Distributed Memory (2014)
12. Lu, T., et al.: A unified storage and query optimization framework for sensor data. In: 2015 12th Web Information System and Application Conference (WISA) (2015). https://doi.org/10.1109/wisa.2015.13. Accessed 26 Jul 2024
13. Mehta, S., et al.: Empirical Evidence of Heuristic and Cost Based Query Optimizations in Relational Databases
14. Niu, X., et al.: Heuristic and cost-based optimization for diverse provenance tasks. IEEE Trans. Knowl. Data Eng. **31**(7), 1267–1280 (2019). https://doi.org/10.1109/tkde.2018.2827074. Accessed 26 Jul 2024
15. Paudel, N., Bhatta, J.: Cost-based query optimization in centralized relational databases. J. Inst. Sci. Technol. **24**(1), 42–46 (2019). https://doi.org/10.3126/jist.v24i1.24627. Accessed 26 Jul 2024
16. Marathe, A., et al.: Integrating the Orca Optimizer into MySQL. https://doi.org/10.48786/edbt.2022.45. Accessed 26 Jul 2024

17. Šušter, I., Ranisavljević, T.: Optimization of MYSQL database. J. Process Manag. New Technol. **11**(1–2), 141–151 (2023), https://doi.org/10.5937/jpmnt11-44471. Accessed 26 Jul 2024
18. Kheradkar, V.V., et al.: Query Optimization in Uncertain and Probabilistic Databases (2023). https://doi.org/10.21203/rs.3.rs-3268445/v1. Accessed 26 Jul 2024

Securing Sign-Up and Sign-In in the Indian Market: Fraud Prevention in CIAM for CRM - A Comprehensive Review and Future Directions

Nitesh Behare[1(✉)], Manesh R. Palav[2], Sharmishthadevi B. Deshmukh[2], Shrikant Waghulkar[3], and Vinod Pal[4]

[1] Balaji Institute of International Business, Sri Balaji University Pune (SBUP), Pune, India
beharenb@gmail.com
[2] Dr. D. Y. Patil Vidyapeeth's, Global Business School and Research Centre, Pune, India
[3] Arihant Education Foundation's Arihant, Institute of Business Management, Pune, India
[4] Marwadi University, Rajkot, India

Abstract. Customer Identity and Access Management (CIAM) has alwsays been critical in digital Customer Relationship Management (CRM) systems specifically in safeguarding customer's identity in diverse digital landscapes like India. This paper presents a comprehensive review of CIAM's role in fraud prevention at the time of sign-up and sign-in processes of a customer within CRM platforms, with reference to the Indian market. It underscores the current practices of business processes, key challenges faced by customer as well as business and effective fraud prevention strategies, comprising the use of Multi-Factor Authentication (MFA), biometric verification and machine learning for real-time threat detection. Furthermore, it also analyzes India specific factors such as rules and regulations, people diversity and geographical and technological developments that define CIAM in the country. Moreover, decentralized identity management using blockchain, emerging trends in AI-driven fraud prevention and implementation of Zero Trust Architecture (ZTA) are also discussed to offer insights into the future of CIAM in the CRM domain. This aim of the study is to guide business to enhance their CIAM strategies to ensure vigorous security, regulatory compliance and seamless customer experiences, predominantly in the evolving digital market of India.

Keywords: Customer Identity and Access Management (CIAM) · Digital Identity Management · Cybersecurity in CRM · CRM Security · Sign-Up Security · Sign-In Security · Zero Trust architecture · Homomorphic Encryption · Quantum-Resistant Algorithms · Predictive Analytics

1 Introduction

Two essential components of the advanced digital business model are customer relationship management (CRM) and *Customer Identity and Access Management* (CIAM). CIAM is a system that protects and identifies consumers across all online resources

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 274–301, 2025.
https://doi.org/10.1007/978-3-031-86305-9_19

and offerings from businesses. It is essential because it serves as facilitator between customers and service providers while also ensuring robust security. CRM in this context refers to customer relationship management, which includes interactions and the use of data to send pertinent communications. Altogether, CIAM model and CRM systems contribute to the frameworks of customer centric businesses safeguarding customer outreach and data management to serve as a core of customer centric businesses.

In the field of CIAM for CRM systems, fraud prevention is a critical challenge. Businesses are gradually digitizing their routine operations and customer interactions which has led to proliferate the risks associated with fraudulent activities. Prominent fraud prevention activities are crucial to protect subtle customer data and diminish financial losses and to endorse consumer trust and loyalty (Podugu, Rayapureddi, & Gupta, 2023). Business can defend several cyber threats and ensure adherence to regulatory requirements by implementing vigorous authentication and monitoring mechanisms. In today's digital business environment, where customer data is a crucial asset, the prominence of integrating rigorous fraud prevention strategies into CIAM frameworks within CRM systems cannot be exaggerated (Rasouli & Valmohammadi, 2020). This paper aims at reviewing and analyzing the CIAM solutions in CRM systems today, with emphasis on approaches used in managing fraud threats in digital identification.

In terms of Indian market, it is necessary that CIAM is thoroughly associated with ensuring secured sign-up and sign-in processes within CRM systems (Singh & Sirohi, 2014). The current form of digitalization in India thus features an evolving and expanding internet-using population coupled with heterogeneity in consumers' buying preferences and different degrees of digital literacy. This is dealt with by CIAM in as much as they provide solutions that suit the culture as well as meeting the legal frameworks of a country.

New and upcoming regulations such as the Personal Data Protection Bill (PDPB) require the use of strict practices in the protection of consumers' personal data and hence, necessitate the use of robust CIAM strategies to comply. CIAM not only assures the customer secure customer authentication and access control but also builds up loyalty by protecting the customer's information from impostors (Shehu, Pinto, & Correia, 2020). In India's dynamic and mobile-driven market, CIAM frameworks are critical in supporting digital interactions regarding consumers' privacy and legal requirements (De & Shukla, 2020) for sustainable business development (Podugu, Rayapureddi, & Gupta, 2023).

When considering the dynamics of CIAM within the CRM systems including mainly the fraud prevention in the Indian market, a number of critical research questions surfaces as follows. These questions explore current CIAM frameworks' efficacy, how regulations affect them and the significant influence of cultural factors on consumers' behaviour, comparative assessment of fraud prevention practices, novel technologies and their implications on global customer trust and loyalty (Hassan, et al., 2024). Thus, answering these questions, this study aims to contribute to the understanding of global CIAM development and outline the strategies for the Indian context given the peculiarities of the region's IT development.

Following are some research questions that could explore CIAM in CRM with a focus on fraud prevention in the Indian market:

- How do current CIAM frameworks in CRM systems mitigate fraud risks during customer sign-up and sign-in processes, specifically in the Indian market?
- What are the regulatory challenges and compliance requirements that influence the implementation of CIAM strategies for fraud prevention in CRM systems in India?
- What are the cultural factors that shape consumer perceptions and expectations regarding CIAM practices in the Indian digital landscape?
- How do different fraud prevention techniques, such as multi-factor authentication and biometric verification, compare in effectiveness and adoption within CRM systems in India?
- What are the emerging trends and technologies in CIAM that could enhance fraud prevention capabilities in CRM systems, particularly in the context of the Indian market?
- What are the implications of CIAM on customer trust and loyalty in Indian businesses, considering the balance between security measures and user experience?

2 Literature Review

2.1 Overview of Existing Research on CIAM

A brief literature review of CIAM demonstrates diverse and complex area with objectives related to improving customers' digital security (Mandal & Amilan, 2023), personalized services and following the regulatory compliances. Many studies have revealed that CIAM plays a major role in handling customer identities throughout multiple digital platforms with the aim of providing reliable accesses to the services furnished while protecting sensitive personal data (Pohn & Hommel, 2023). Key themes in CIAM research embraces authentication methods followed by authorization frameworks and identity lifecycle management (Rasouli & Valmohammadi, 2020). Some studies have explored the diverse authentication mechanisms based on password, multi-factor authentication (MFA) and/or biometric verification, gauging their effectiveness in balancing security with usability (Aleluya & Vicente, 2018), (Karim, et al., 2024). The evolution of CIAM frameworks incorporates adaptive authentication models that vigorously adjust security measures on the basis of contextual elements like type of device, location as well as user behaviour (Bumiller, et al., 2022).

Many studies have also explored the regulatory environment that affects implementation of effective CIAM model with important considerations of data protection regulations including GDPR, CCPA and customs of other countries (Poritskiy, Oliveira, & Almeida, 2019). These regulations require proper data processing controls, clearly stated and explained data processing operations and a proper approach to managing users' consent (Drozd & Kirrane, 2020) which define CIAM solutions worldwide (Viguri Cordero, 2021). Currently, developments in the field of CIAM involve improving the detection of fraudulent activities with the help of integrated analytics, machine learning and monitoring of users' actions in real-time (Tamanna, Kamboj, Singh, & Kaur, 2024). Also, there is an emerging trend to incorporate CIAM with AI for predicting and the preventing identity-related threats at an early stage (Ludera, 2021).

2.2 Fraud Prevention Techniques in Digital Identity Management

The methodological approaches which are followed in digital identity management for fraud prevention are wide-ranging practices that are applied to protect the individual's sensitive data from misuse (Tripathi & Tripathi, 2024) (Soomro, Ahmed, Shah, & Khoumbati, 2019). These techniques are important in relieving different types of identity fraud such as account takeover frauds, phishing, identity theft and other related scams that can result into loss of money as well as reputational damage for individuals and similarly for companies (Bhatt, Gupta, & Sharman, 2016). One straightforward method of fraud prevention is Multi-Factor Authentication (MFA) which is a security protocol that makes the users give authentication details of not less than two different characteristics like passwords, biometric features such as fingerprints or face recognition, smart cards and special one-time passwords (OTP) (Taher, Nahar, & Hossain, 2019). MFA also increases the protection of information by supplementing the regular password that a user is likely to have a hard time protecting in the event the other factors are penetrated.

Biometric identification has proved to be a very secure and effective method of fraud prevention using physical traits such as fingerprints, iris or facial characteristics to approve the users' identity (Carmel & D., 2020). Biometrics scores high on the accuracy level and are hard to imitate or forge and hence, useful in curbing cases of identity fraud. Another advanced method is Behavioural analytics that tracks the activity of users and their behaviour to identify things that point to fraud. Behavioural analytics collects and analyzes data about users' login times, device locations and transaction histories to spot the suspicious activities and raise security alerts to address activities that are flagged as suspicious in real time.

In addition, fraud prevention strategies enlist the use of machine learning and artificial intelligence (AI) to analyze large data set constantly to detect new forms of the fraud (Xu, Wang, Liao, & Wang, 2023). These technologies help in the prevention of fraud since the methods learn how to prevent new threats in a step by step pattern thus improving the security of digital identity management systems. (Vivek, Nandhan, Zean, D., & Dhanwanth, 2023). Together, these techniques comprise the holistic approach to secure digital identities and insightful information, thus building the trust among the users and providing the secured access to the digital services and assets.

2.2.1 CIAM Implementations in the CRM Domain

CIAM solutions in the CRM space concentrate on incorporating reliable identity protection services together with customer relationship initiatives. The purpose of these implementations is to simplify processes of user identification across various points of contact with a user, in order to provide the user with the convenience of secure access to CRM systems and services (Rasouli & Valmohammadi, 2020). Maintaining centralized customers' identity repository that allows distinctive communications to enhance dealings with it as well as increase effectiveness are the important characteristics of these systems. In CRM, CIAM solutions also focus on meeting the requirement on data protection acts as well as keeping the users' privacy and consent a priority (Bumiller, et al., 2022), (Kobielus, 2023). The use of CIAM solutions contributes not only to the strengthening of the security system but also to the further development of users' confidence and appreciation of the services provided. Through the integration of CIAM with

CRM systems, customer engagement strategies can be enhanced while protecting the customers' personal data, hence helping organisations in achieving their business goals successfully (Markkanen & Frantti, 2023).

2.2.2 Specific Challenges and Solutions in the Indian Market

The Indian market has its own challenges predominantly with reference to digital security and user trust services. For implementation of Customer Identity and Access Management (CIAM) strategies in the Indian market need identifying the associated key issues (Kulkarni & K.B., 2019). A recent alarming trend has underscored serious need for robust digital security measures in CIAM systems. In India, some fraudsters have started stealing customer's data such as thumb impressions to create duplicate prints on silicon, from Haryana's land record's website, which was used to withdraw money from Aadhaar-linked bank accounts of credulous victims. This prompted Faridabad Police to endorse an audit of official website and app used to access the state's land record (Pavneet Singh, 2022).

Diverse demographic and cultural landscape of India is one of the significant challenge in implementing CIAM. Varying levels of digital literacy and differing technology adoption rates across demographic segments considerably influence user behaviours and expectations related to data privacy and security. This diversity necessitates tailored approaches to CIAM to effectively cater to different user groups while ensuring robust security.

The second big challenge is regulatory compliance because of the complexity in today's regulatory environment consisting of many laws and rules as well as frequently changing rules and regulations. Organization has to deal with laws such as the Personal Data Protection Bill (PDPB) and other rules and regulations issued by the Reserve Bank of India (RBI) and other regulatory authorities that can be extremely uncomfortable (Anish, 2024). These regulations impose rigorous data protection measures, like consent management, data localization and timely breach notifications. Adherence with these instructions is essential in designing and implementing effective CIAM frameworks.

Additionally, the mobile Internet is gaining more popularity in India and this requires mobile-friendly CIAM solutions that are compatible with different kinds of mobile devices and networks. This calls for proper changes in the process of user authentication and the user interfaces considering choices and restrictions because of the mobility of the devices (Yadav, Mishra, & Prakash, 2013). The socio economic factors introduces potential challenges in ensuring consistent and secure digital environment as availability of internet and other related solutions differ across urban and rural areas (Bensari & Bilami, 2013). Dealing with these challenges demands scalable, flexible and cost effective CIAM solutions for the Indian population's diverse segments to ensure accessibility and usability (Freeman & Park, 2015).

3 Current State of CIAM in CRM

Customer Identity and Access Management (CIAM) as part of the Customer Relationship Management (CRM) system is concerned with the handling and protecting customer identities at the multi-channel interface. This section gives the outline of current scenario of the CIAM practices in CRM, underlines the mainstream security threats and fraud threats along with the analysis of case studies from the Indian market.

3.1 Overview of Current CIAM Practices in CRM Systems

CIAM in CRM systems also entails the application of techniques and tools for identifying and authorizing the users, managing their identities and controlling access to the various features of the CRM. Key practices include:

- Centralized Identity Management: Generally, CIAM solutions are integrated with CRM systems where centralized customer's identity data can be managed and kept coherent as well as correct across different channels or interfaces (Cabinakova, Ostern, & Krönung, 2019).
- Authentication Methods: Various authentication methods are employed that ranges from traditional password-based approaches to advanced techniques such as multi-factor authentication (MFA) and biometric verification. MFA enhances security by necessitating users to offer multiple forms of verification, minimizing the risk of unauthorized access (Charan Ayineni, Pranav, & Vasanth, 2024).
- Single Sign-On (SSO) Integration: SSO capabilities streamline user access by allowing customers to authenticate and gain access to multiple CRM applications and services without needing to re-enter credentials (Almeida, Zúquete, Pazos, & Oliveira, 2024).
- Consent Management: Well-designed CIAM framework in CRM systems include consent management features that allows organizations to determine and manage users' consent on data processing with regards to GDPR and other local data protection regulations (Tokas & Owe, 2020).

3.2 Common Security Challenges and Fraud Risks

Despite advancements in CIAM technology, CRM systems face several security challenges and fraud risks:

- Phishing and Social Engineering: Phishing emails and social engineering tactics used by cyber attackers to obtain login credentials and compromise user accounts in order to take advantage of human weakness (Kenneth, et al., 2023).
- Account Takeover (ATO): These involve getting unauthorized access to user accounts, usually by using poor authentication procedures or credentials that have been stolen (Kawase, Diana, Czeladka, Schüler, & Faust, 2019).
- Data Breaches: When sensitive customer information like personal and financial data, is accessed by malicious actors, can leads to vulnerabilities in CRM systems (Al Kinoon, Omar, Mohaisen, & Mohaisen, 2021).

- Regulatory Compliance: Ensuring adherence to data protection regulations poses serious challenges, predominantly in regions with stringent requirements like India, where stringent laws like PDPB require the implementation of user consent management for robust data security (Maurya & Prasad, 2020).
- Synthetic Identity Fraud: Many times cybercriminals create synthetic identities with the help of combination of actual and fake information, making it challenging for CIAM systems to detect them. This form of identity theft is hard to catch due to the blending of legitimate data with fabricated details (Khan, Bokhari, Hanafi, & Zeyauddin, 2023).
- Evolving Cyber Threats: Fraudsters constantly adapt their techniques, making it difficult for static CIAM systems to keep up. Continuous updates to threat models are required to address novel attacks.
- Balancing User Experience and Security: Maintaining right balance between mechanisms of resilient authentication and unified user experiences remains a challenge. Excessive security checks can drive customers away, while too few can lead to vulnerabilities (Karim, et al., 2024).

3.3 Case Studies from the Indian Market

3.3.1 Case Study 1: Uber's Sign-Up Fraud Incident

In 2017, Uber, the global ride-sharing firm experienced a major sign-up fraud outrage that exposed flaws in regard to their user identification process. Scammers made use of vulnerabilities in their system by forging new accounts using fake identities. Such scams incorporate themselves to take benefit out of the bonus and rebates leading to the financial loss to the company (Julia, 2017).

Uber countered by strengthening its CIAM architecture bringing in better and more rigid identity check mechanisms as well as better fraud identification techniques. They put measures such as Multi-factor authentication (MFA) and improved mode anomaly detection to prevent the creation of new accounts with malicious intention immediately. All these measures had the effect of not only preventing future fraud occurrences but also in reinstating the confidence of the customer through secure and reliable provision of services.

3.3.2 Case Study 2: Yahoo's Massive Data Breach

Yahoo.com was hacked twice once in 2013 and second time in 2014 where over 3 billion accounts compromised in the largest data breach in the company's history. This act of infringement involved invading the organisation's network, extracting account information and credentials. Attackers targeted vulnerabilities in sign-in mechanisms of Yahoo making them access user information such as names, email addresses, phone numbers and hashed passwords. This incident underscored the need of CIAM to have strong security measures in place concerning authentication and data protection. Yahoo reacted by implementing stronger password policies, augmenting encryption protocols and introducing account security features like two-step verification. These measures aimed to prevent unauthorized sign-ins and protect user data from future breaches (Online, 2020).

3.3.3 Case Study 3: Facebook and Cambridge Analytica Scandal

Following the Cambridge Analytica scandal, which exposed the unapproved harvesting of millions of Facebook users' personal data, Facebook came under heavy criticism in 2018. The incident concerned a third-party app that took advantage of holes in Facebook's data sharing policies to gather user data through the sign-in feature (Carole & Emma, 2018). The CIAM practices of Facebook came under the scan for their inability to provide proper security to the data collected from users and implement stringent user consent as a measure in devices. In response, Facebook tightened its API regulations, introduced improved privacy settings to its users and investigated the access of third party applications in a bid to curb rogue data access and improve on user credentialing (IGA, 2018).

3.3.4 Case Study 4: Marriott International's Guest Reservation System Breach

In the year 2018, Marriott International revealed largest data breach that compromised about 500 million guest records. This breach entailed hacking of Marriott's guest reservation database and acquired personal information including their name, contact details, passport numbers and even their encrypted payment card data of many guests (BBC, 2018). This incident revealed the security gaps in Marriott's CIAM and CRM systems specifically in sing-in processes and data protection. As a result, Marriott enhanced encryption of data, required MFA for its employee access, especially for data-sensitive functions and strengthened monitoring and analytics for timely identification of unauthorized access attempts and respond accordingly (Sujith, 2024) (Li, 2023).

These cases highlighted the variations of sign-up and sign-in fraud that occurs in various industries and prove significance of strong CIAM frameworks to prevent such risks and protect customers' data. In a view of executing successful fraud prevention steps like robust authentication elements, data encryption and constant monitoring should remain a priority for all organizations to ensure the safety of user identities and the given platforms' credibility.

4 CIAM and CRM Integration in the Indian Market

4.1 Specific Considerations for the Indian Market

In the Indian market, integrating CIAM with CRM systems requires a nuanced approach personalized to the country's unique dynamics., With its vast population and speedily expanding digital ecosystem, India presents both opportunities and critical challenges for businesses that are aiming to secure customer identities and elevate customer relations.

A major factor that is pertinent to look at is heterogeneity of Indian population in terms of level of internet penetration and their proficiency. Whereas, large number of internet users and tech-savvy customer base can be targeted in urban locations, rural users may have very restricted access to such digital services and hence may be comparatively less digitally literate (Freeman & Park, 2015). Hence, CIAM solutions should be natively integrated and adoptable across devices types and internet connectivity scenario. Simplification for easy use while maintaining security is paramount for unlocking usage on more platforms (Freeman, Park, & Middleton, 2020).

Another crucial factor is indeed the high mobile usage in India where large number of users belongs to. Given that millions of mobile users engage mostly with services, CIAM solutions should be tailored to mobile interfaces. This includes mobile-friendly measures for authentication such as biometric identification and one time passwords(OTPs) which are easily recognizable and suitable for users.

4.2 Regulatory and Compliance Requirements

The regulatory landscape in India plays a key role in the CIAM and CRM integration process which oblige organizations to follow extensive measures to safeguard customer data and ensure user anonymity. Key regulations and guidelines determine the designing and implementation framework to meet some specific standards and increase consumers' confidence level.

4.2.1 Personal Data Protection Bill (PDPB)

The Personal Data Protection Bill (PDPB) is a central law in India that seeks to protect the personal data and further strengthen the rights to privacy for individuals (Maurya & Prasad, 2020). When managing customer identities and data, the PDPB mandates numerous rigorous measures that an organization must adhere to. Key provisions include:

- User Consent: Organizations are prohibited from collection, processing or sharing of users' personal data without their permission. This requires the assurance of users and subjects about the reasons for data collection and usage of the collected data (Tokas & Owe, 2020).
- Data Localization: The PDPB makes some specific localisation obligations concerning personal data, especially sensitive data, in India. This implies that organizations must have adequate infrastructure to comply with the localization conditions (Vasishtan, 2022), (Fan, 2023).
- Transparency in Data Processing: A key component of accountability is a fair processing of data by the organizations, the users of the data must be informed about data processing activities and data processing has to be done fairly and lawfully (Steenbergen & Meesters, 2021).

To meet the guidelines set by the PDPB, it is mandatory to have advanced CIAM solutions that focus on effective data protection for users' data and user's privacy. This comprises of authorizing individuals' access in a secure manner, using encryptions to protect sensitive data and making sure that the ways of processing data are transparent to its users and the governing regulations.

4.2.2 Reserve Bank of India (RBI) Guidelines

The Reserve Bank of India (RBI) has release special norms for financial institutions to safeguard customer information and prevention of fraud. These guidelines call for rigorous policies for authentication and protection of sensitive financial data. Key aspects of the RBI guidelines include:

- Multi-Factor Authentication (MFA): In order to increase the level of protection, financial institutions are obliged to use MFA. MFA entails several forms of identification process like passwords, biometric and OTPs to ensure that sensitive financial information should only be accessed by the authorised users only (Karim, et al., 2024).
- Encryption: According to RBI guidelines it is compulsory to encrypt the data when it is transmitted and when it is stored in the database. This warrants that even if the data is intercepted or gets into the wrong hands, it remains secure as it can only be read by an individual controller (Gupta & Jain, 2018).

FinTech organizations have to implement CIAM solutions that meet these requirements and enhance the security of customers' identities and their financial data. The two main common eventualities that are also mandatory include security audits and security assessments which are needed for confirmation of continued compliance and the detection of susceptibilities.

4.2.3 Sector-Specific Regulations

Other regulations followed in this context in addition to the PDPB and RBI guidelines are sector specific standards in the telecommunication and healthcare industry that demands the protection of the customer data. These regulations impose additional requirements on organizations operating in these sectors:

- Telecom Regulatory Authority of India (TRAI): The TRAI focuses on the important aspects of security measures in access to customer data in telecom services. Telecom providers require strict access controls policies where only the people who are permitted are allowed to access such information. Regular audits and compliance checks are must to ensure that all rules and regulations put forth by TRAI are being met (Arun, 2018).
- Healthcare Sector: The applicable health care standards require the safeguarding of patients' sensitive health data. Healthcare provider must adopt CIAM solutions that meet the requirements of protecting patient data and only allow access to pertinent staff. This encompasses the practice of protection of data by use of data encryption, robust user authentication methods and periodic security audits (Jawad, 2024).

CIAM compliance entails several sector-specific standards to align with industry guidelines implementing stringent access control, data encryption, and frequent compliance audit. Thus, following regulatory requirements support organizations in improving data protection, customers' privacy and trust among the users.

4.3 Cultural and Technological Factors

4.3.1 Cultural Factors

Culture elements play a crucial role in determining awareness and acceptance of CIAM and CRM systems in India. The insight of these cultural differences is vital for the organizations which seek to establish trustful and lasting relations with their consumers (Milisavljević, et al., 2013).

4.3.2 Trust and Personal Relationships

- High Value on Trust: Indians are very particular in respecting the trust and building a healthy personal relation. This cultural characteristic also applies to their communication with digital services since they use those platforms they trust (Simonds, Goins, Krantz, & Garroutte, 2014).
- Building Trust: In addition, organizations have to ensure the transparency of data processing and proper identity management as the key aspects of consumers' trust. This involves clearly communicating how customer data is collected, used, and protected. Users need to feel confident that their personal information is handled with utmost care and security (Steenbergen & Meesters, 2021).

4.3.3 Communication and Language

- Diverse Linguistic Landscape: India has a rich linguistic culture with many languages and dialects. Effective multilingual communication facilitates constructive user involvement and in turn, improves users' trust.
- Localized Communication: Localized communication strategies through multilingual support should be adopted by organizations while implementing CIAM and CRM systems to address the linguistic differences in Indian population.
- Technological Factors: Further, technological innovations and the distinct style of the Indian IT market also come into the picture while modelling CIAM and CRM integration.

4.3.4 Advancements in Digital Infrastructure

- 4G and 5G Adoption: India has seen a tremendous growth of digital infrastructure with most people using 4G and soon they will move to the next level of 5G networks. Such developments make it easier to integrate and achieve a higher bandwidth, which in turn, power more improved CIAM solutions (Kalyan, 2024).
- Enhanced CIAM Capabilities: Better connectivity provides organisations with an opportunity to adopt real time fraud detection system and advanced analytics powered by Artificial Intelligence to improve on security and the user experience. Faster networks also have an added advantage in ensuring smooth and secure authentication process creating a seamless and satisfying customer's interaction experience (Aabitova, et al., 2024).

4.3.5 Fragmented Technological Landscape

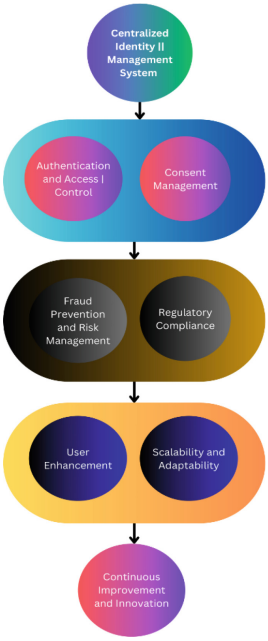
- Varying Device Capabilities: Technological infrastructure of India is not very coherent at the present, as standards of devices and network reliability differ. It targets users from those who own least a basic feature phone to those who possess advanced smart phone (Shaik, Borgaonkar, Park, & Seifert, 2019).
- Adaptable CIAM Solutions: In these instances, CIAM solutions must remain effective and provide same benefits maintaining the integrity across all environments. This includes the enhancement in the authentication methods suitable for the limited bandwidth environment as well as compatibility with a number of devices (Memon, 2015). For example, using easy mobile authentication methods such as based on SMS and OTP for poor network coverage environments.

4.3.6 Integration of Emerging Technologies

- AI and Machine Learning (ML): The modern technologies like AI and ML can greatly improve CIAM systems. These technologies foster prevention of possible fraud threat through its predictive risk analysis mechanism. Analysing users' behaviour patterns, AI can foresee deviations and activate the security controls (Raudha & Saeedi, 2019).
- Personalized User Experiences: It is also useful in customizing the experience of the users in CRM systems by studying their behaviour patterns and preferences. Such analysis enables the organizations to deliver messages that make a favourable impact on the Indian consumers and increase the customer satisfaction and trustworthiness (Tzafilkou, Protogeros, & Yakinthos, 2015). For example, unique advertising and other marketing communications may be delivered based on users' profile information and their activity.

4.3.7 Digital Literacy and Accessibility

- Varying Levels of Digital Literacy: Urban areas in India are already tech savvy; the digital literacy of any rural areas may still be low. CIAM systems should be easy to understand and convenient to interact regardless of the users' levels of digital literacy (Nedungadi, Menon, Gutjahr, Erickson, & Raman, 2018).
- Accessibility Features: Some of the general design principles include supporting voice based controls and easier and less cluttered user interface can make digital service comprehensive and user-friendly (Vendome, Solano, Linan, & Linares-Vasquez, 2019).
- CIAM Model



5 Future Directions

- Emerging trends in CIAM and fraud prevention

5.1 Advanced Authentication Mechanisms

Biometric Authentication:

- Increased Adoption: Biometric authentication methods such as fingerprint scanning, facial recognition, and voice recognition are increasingly being adopted. These methods provide higher security as compared to traditional password-based systems (Ibrahim, Tamimi, & Abdalla, 2017).
- Seamless User Experience: Biometrics offer a seamless and convenient user experience, reducing friction during the authentication process and improving user satisfaction (Wiercioch, Teufel, & Teufel, 2018).

Multi-factor Authentication (MFA):

- Enhanced Security: MFA, which combines multiple verification methods (e.g., something you know, something you have, something you are), is becoming standard practice. This significantly enhances security by adding layers of protection (Simha.R, H.K., Prabhu, & Joshi, 2024).
- Adaptive Authentication: Adaptive MFA adjusts the authentication requirements based on the user's behaviour and risk profile, providing an optimal balance between security and user convenience (Pramila, Misbahuddin, & Shukla, 2022).

5.2 Artificial Intelligence (AI) and Machine Learning (ML)

AI-Powered Fraud Detection:

- Predictive Analytics: AI and ML are used to analyze vast amounts of data to identify patterns indicative of fraudulent activity. Predictive analytics can detect and mitigate fraud in real-time, even before it occurs (Soviany, 2018).
- Behavioural Biometrics: AI analyzes user's behaviour, such as typing speed, mouse movements, and navigation patterns, to create unique behavioural profiles. Deviations from these profiles can trigger security alerts (Bhatt & T., 2013).

Continuous Authentication:

- Ongoing Verification: Continuous authentication uses AI to monitor user activities throughout a session, ensuring that the person who logged in remains the same. This process of ongoing verification helps in discovering user's session hijacking and other fraudulent activities (Barone & Mekni, 2023).

5.3 Decentralized Identity (Self-Sovereign Identity)

Control Over Personal Data:

- User Empowerment: Decentralized identity (also known as Self-Sovereign Identity) give rights to the individuals to control their personal data and empower them to manage and share their identity information without depending on centralized entities (Thorve, Shirole, Jain, Santhumayor, & Sarode, 2022).
- Blockchain Technology: Blockchain technology always used to implement decentralized identity solutions to safeguard data integrity as well as security through distributed ledger technology (Barone & Mekni, 2023).

Privacy and Security:

- Enhanced Privacy: Decentralized identity systems mitigate risks related with breach of data and unauthorized access by minimizing dependency on centralized databases.
- Interoperability: These systems not only encourage interoperability among various platforms and services but also simplify identity verification processes across numerous contexts.

5.4 Zero Trust Architecture

Trust No One Model:

- Comprehensive Security: The Zero Trust Model (ZTM) considers that threats can exist in the network i.e. both inside and outside of the network. Strict verification for every user and device attempting to access the resources are very much needed (Azad, Abdullah, Arshad, Lallie, & Ahmed, 2024).

- Micro-Segmentation: Micro segmentation technique encompasses segmentation of the network into smaller zones where each zone will have its own security policies. Admission to these zones is strongly controlled and monitored by the system (Dhiman, et al., 2024).

Continuous Monitoring:

- Real-Time Analysis: User activities and network traffic can be continuously monitored to detect and respond to the threats in real-time to ensure vigorous security (Barone & Mekni, 2023).

5.5 Privacy-Enhancing Technologies

Data Minimization:

- Reducing Data Collection: 'Privacy-enhancing technologies' emphasises on reducing the amount of data collected and analysed to reduce risk of exposure and misuse of the user's personal data.
- Anonymization and Pseudonymization: Anonymization and Pseudonymization approaches are used to safeguard user's personal data and making it challenging to trace back to individuals (Bolognini & Bistolfi, 2017).

Homomorphic Encryption:

- Secure Computations: Homomorphic encryption allows calculations to be executed on encrypted data without decrypting it. Homomorphic encryption ensures data privacy while empowering useful analysis and processing (Jain, Cherukuri, & Kamalov, 2024).

5.6 Regulatory Compliance and Data Protection

Across the globe, regulatory compliance expects rigorous data protection methods which includes consent management, breach notifications and data localization, safeguarding secure and privacy-respecting CIAM implementations within multifaceted legitimate frameworks.

GDPR and Beyond:

- Stringent Regulations: Regulatory compliance like California Consumer Privacy Act (CCPA), General Data Protection Regulation (GDPR) of European Union, Personal Data Protection Bill(PDPB) of India and other data protection laws is driving the adoption of robust CIAM solutions (Labadie & Legner, 2020).
- Global Standards: Gradually all the business are aligning their data protection standards with global standards to certify compliance and build customer trust.

Data Residency and Localization:

- Local Compliance: Data residency and localization requirements mandate that personal data be stored within specific geographical boundaries. This trend is influencing CIAM strategies to ensure compliance with local laws.

6 Potential Advancements in Technology and Methodology

6.1 Quantum Computing and Cryptography

Quantum-Resistant Algorithms:

- Enhanced Security: As quantum computing advances, there is a growing need for quantum-resistant cryptographic algorithms to protect data against potential quantum attacks (Szatmáry, 2024).
- Future-Proof Security: Developing and implementing these algorithms will ensure long-term data security, even in the face of emerging quantum threats (Yalamuri, Honnavalli, & Eswaran, 2022).

Quantum Key Distribution (QKD):

- Secure Communication: QKD leverages the principles of quantum mechanics to enable secure communication channels, providing theoretically unbreakable encryption keys (Khan, Meraj, & Khan, 2020).

6.2 Blockchain Technology

Decentralized Identity Management:

- Self-Sovereign Identity (SSI): Blockchain can facilitate decentralized identity management systems where users have full control over their personal data, enhancing privacy and security (Naik, Grace, & Jenkins, 2021).
- Immutable Records: Blockchain's immutable ledger ensures the integrity and security of identity data, reducing the risk of data breaches and fraud.

Smart Contracts:

- Automated Processes: Smart contracts on blockchain can automate identity verification and access control processes, ensuring transparency and reducing the potential for human error or fraud (Arun, 2018).

6.3 Artificial Intelligence (AI) and Machine Learning (ML)

Advanced Fraud Detection:

- Behavioural Analytics: AI and ML can analyze user behaviour patterns in real-time to detect anomalies and potential fraud activities more accurately (Bhatt & T., 2013).
- Adaptive Security Measures: These technologies enable adaptive security measures that respond dynamically to detected threats, enhancing overall security.

Personalized User Experiences:

- Tailored Interactions: AI-driven analytics can personalize user interactions based on behavioural data, improving customer satisfaction and loyalty in CRM systems.

6.4 Biometric Innovations

Multimodal Biometrics:

- Enhanced Accuracy: Combining multiple biometric modalities (e.g., fingerprint, facial recognition, iris scanning) can improve the accuracy and security of authentication processes (Samatha & Madhavi, 2024).
- User Convenience: Multimodal biometrics offer a more seamless and user-friendly authentication experience.

Behavioural Biometrics:

- Continuous Authentication: Analysing unique user behaviours (e.g., typing patterns, gait) can enable continuous authentication, reducing the risk of session hijacking and unauthorized access (Senarath, et al., 2023).

6.5 Internet of Things (IoT) Security

IoT Identity Management:

- Secure IoT Networks: Developing CIAM solutions tailored for IoT environments to manage identities and access controls for connected devices, ensuring secure interactions (Deborah, Rajalakshmi, & Mirnalinee, 2024).
- Edge Computing: Implementing edge computing can enhance the security and efficiency of IoT networks by processing data closer to the source.

Device Fingerprinting:

- Enhanced Verification: Device fingerprinting techniques can uniquely identify and authenticate IoT devices, enhancing security by preventing unauthorized device access.

6.6 Privacy-Enhancing Technologies

Federated Learning:

- Data Privacy: Federated learning enables AI models to be trained across decentralized devices without sharing raw data, preserving user privacy while still benefiting from collective learning (Yan, et al., 2020).
- Improved Security: This approach reduces the risk of data breaches by keeping sensitive data localized on user devices.

Homomorphic Encryption:

- Secure Data Processing: Homomorphic encryption allows computations to be performed on encrypted data without decrypting it, ensuring data privacy during processing (Jain, Cherukuri, & Kamalov, 2024).
- Enhanced Compliance: This technology can help organizations comply with stringent data protection regulations by ensuring data remains secure throughout its lifecycle.

6.7 Zero Trust Architecture

Micro-Segmentation:

- Granular Access Control: Implementing micro-segmentation divides the network into smaller, secure segments, each with its own access controls, reducing the attack surface (Basta, Ikram, Kaafar, & Walker, 2022).
- Enhanced Monitoring: Continuous monitoring and analysis within each segment can detect and respond to threats more effectively (Basta, Ikram, Kaafar, & Walker, 2022).

Identity-Centric Security:

- Dynamic Policies: Zero Trust architecture emphasizes identity-centric security, where access policies are dynamically adjusted based on real-time risk assessments and user behaviour (Basta, Ikram, Kaafar, & Walker, 2022).

7 Proposes CIAM Model

7.1 Detailed Explanation of Proposed CIAM Model:

i. User Interaction

Entry Point: The user interaction phase is the crucial entry point where customers engage with CIAM (Customer Identity and Access Management) and CRM (Customer Relationship Management) systems. This encompasses the sign-up and sign-in processes, where users create accounts and authenticate their identities. Ensuring a smooth

and secure user experience during these interactions is vital for building trust and encouraging user engagement.

ii. Trust and Transparency

Building Trust: Trust and transparency are fundamental to successful CIAM and CRM integration. Businesses must prioritize transparent data practices, clearly communicating how customer data is used and protected. This includes obtaining explicit user consent and ensuring users understand the security measures in place. By adopting robust identity management practices, businesses can enhance user trust and foster long-term customer relationships.

iii. Multilingual Communication

Effective Communication: Given India's linguistic diversity, supporting multiple languages in user interfaces and communication materials is essential. Multilingual communication ensures that customers from different regions can understand and engage with the services offered, thereby broadening the reach and improving user satisfaction.

iv. Advanced Authentication

Enhancing Security: Utilizing advanced authentication methods such as biometrics (e.g., fingerprint, facial recognition) and multi-factor authentication (MFA) is crucial for enhancing security. These technologies add multiple layers of protection, making it more difficult for unauthorized users to gain access to sensitive information, thereby reducing the risk of fraud.

v. AI and ML

Fraud Detection and Personalization: Artificial Intelligence (AI) and Machine Learning (ML) play a significant role in CIAM and CRM systems. AI-driven analytics can detect fraudulent activities in real-time by analyzing user behavior patterns and identifying anomalies. Additionally, AI and ML can personalize user experiences by tailoring interactions based on individual preferences and behaviors, thereby enhancing customer satisfaction and loyalty.

vi. IoT Security Management

Secure Interactions: With the proliferation of Internet of Things (IoT) devices, managing identities and access controls for these devices is becoming increasingly important. CIAM solutions must ensure secure interactions between IoT devices and systems, protecting against unauthorized access and potential security breaches.

vii. Regulatory Compliance

Adhering to Regulations: Compliance with local regulations, such as the Personal Data Protection Bill (PDPB), Reserve Bank of India (RBI) guidelines, and sector-specific mandates from bodies like the Telecom Regulatory Authority of India (TRAI), is essential. Adhering to these regulations ensures data security and user privacy, and helps businesses avoid legal repercussions and build trust with consumers.

viii. Emerging Technologies

Future-Proof Security: Integrating cutting-edge technologies like quantum-resistant algorithms and blockchain can future-proof security measures. Quantum-resistant cryptographic algorithms protect against potential quantum computing threats, while blockchain technology facilitates decentralized identity management, ensuring data integrity and user control over personal information.

ix. Zero Trust Architecture

Comprehensive Threat Mitigation: Implementing a Zero Trust security model is critical for comprehensive threat mitigation. Zero Trust assumes that threats can come from both inside and outside the network. It involves continuous monitoring, micro-segmentation to create isolated network segments, and dynamic access control policies that adjust based on real-time risk assessments. This approach minimizes the attack surface and enhances overall security.

By focusing on these key components, businesses can create a robust and effective CIAM and CRM integration model tailored to the Indian market. This model not only addresses cultural and technological factors but also incorporates advanced technologies and regulatory compliance, ensuring secure, efficient, and user-friendly interactions. This comprehensive approach enhances security, fosters trust, and improves customer engagement, positioning businesses for long-term success in India's dynamic digital landscape.

7.2 Strategic Directions for Businesses in the Indian Market

7.2.1 Building Trust and Leveraging Advanced Technologies

The Indian Market is one of the most diverse and is rapidly shifting to the digital form of doing business can be deemed as the opportunity as well as the threat for the business. Trust creation and management as well as appropriate use of advanced technologies are yet other strategic mechanisms worthy of consideration. Trust and transparency are significant; businesses must prioritize data privacy by adopting transparent data practices, clearly communicating how customer data is used and protected (Ameyed, Jaafar, Charette-Migneault, & Cheriet, 2021). Therefore, it can be deduced that compliance to the provisions of the PDPB continues as key to gain credibility and trust among consumers. Furthermore, building customer relationships through personalized communications is important.

The CRM systems should be used to track the customer preferences and behaviours that would help the businesses in formulating and employing marketing strategies and services appropriate to the customers. Applying intelligent technologies and tools such as Artificial Intelligence and Machine Learning can fully transform the sphere of fraudulent actions and customer interaction. Critical application of AI can result in fraud detection in real-time (Aabitova, et al., 2024) and provide tailored customers' interactions and experiences based on patterns of behaviour and fraud detection. Similarly, the integration of biometric and multi-factor authentication (Karim, et al., 2024) enhances the security and ease of use from the side of the users. It is also recommended that businesses integrating IoT for the better management of identity across the internet and blockchain for its decentralized characteristics that help in the maintenance of data consistency as well as improve security protocols.

7.2.2 Technological Adaptation and Regulatory Compliance

It is critical for the success of business that one should adapt diverse technological landscape. Businesses must optimize their CIAM solutions to perform well across varying connectivity environments, from low-bandwidth areas to advanced urban centers (Purwanto & Soewito, 2021). Ensuring compatibility with a wide range of devices, from basic feature phones to high-end smartphones, is necessary. As India progresses towards widespread 5G adoption, businesses should prepare to leverage enhanced connectivity benefits, such as real-time data processing, improved mobile experiences, and faster service delivery. Staying ahead of regulatory changes is also crucial. Local legislation that includes but not limited to PDPB must be complied with as well as other sectorial regulation guides from Reserve Bank of India (RBI) and Telecom Regulatory Authority of India (TRAI). Companies should set up specialist groups to track the possible alterations in the laws and regulations and compliance and being constantly up to date. Using measures like Zero Trust that are predicated on the idea that threats can be from internal and external sources will aid in protection of sensitive data. This includes constant vigilance, splitting a large subnet into small portions and dynamic access control policies. Furthermore, digital literacy trainings and adopting interface enhancements like voice-over and simplified designs will enhance accessibility to internet throughout numerous population segments. Based on the above-stated strategic directions, businesses can easily work through complexities of the Indian market improving security, user interfaces as well as nurturing customer bases.

8 Conclusion

To successfully operate in Indian market, companies must incorporate trust-creation efforts, cutting-edge technologies, flexibility to the variety in the technological environments, compliance with Indian regulations and investments for digital competency. Focus on establishing trust and transparency through sound data protection and individualized approach to customers helps to established strong and lasting relationship and hence are critical. Integration of new technologies like AI, face recognition, biometrics and block chain improves security as well as add value for improving user experience. By aligning with India's differential connectivity and readiness for 5G, businesses maintain its relevance to competitive launches of effective and prompt services. Compliance with the legislative measures such as the PDPB along with sectorial standards results in legal compliance and boosts customer confidence. Adopting Zero-Trust security model and improving digital inclusiveness append a stronger foundation for business proceedings in the continuously growing digital environment. Thus, it can be stated that by following these strategic directions, businesses can successfully manage threat and seize opportunities and build long-term positive relationship with Indian consumers.

References

Aabitova, G., Abalkanov, M., Abitova, G.A., Shuteyeva, G., Aitmukhanbetova, E., Kulniyazova, K.: Review of cloud AI for real-time fraud detection. In: 0th International Conference on Automation, Robotics, and Applications, ICARA 2024, Athens, pp. 454–460. Institute of

Electrical and Electronics Engineers Inc. (2024). https://doi.org/10.1109/ICARA60736.2024.10553149

Al Kinoon, M., Omar, M., Mohaisen, M., Mohaisen, D.: Security breaches in the healthcare domain: a spatiotemporal analysis. In: Mohaisen, D., Jin, R. (eds.) Computational Data and Social Networks. LNCS, vol. 13116, pp. 171–183. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-91434-9_16

Aleluya, E.R., Vicente, C.T.: Faceture ID: face and hand gesture multi-factor authentication using deep learning. In: 3rd International Conference on Computer Science and Computational Intelligence, ICCSCI 2018, Tangerang, vol. 135, pp. 147–154. Elsevier B.V. (2018). https://doi.org/10.1016/j.procs.2018.08.160

Almeida, J.R., Zúquete, A., Pazos, A., Oliveira, J.L.: A federated authentication schema among multiple identity providers. Heliyon **10**(07) (2024). https://doi.org/10.1016/j.heliyon.2024.e28560

Ameyed, D., Jaafar, F., Charette-Migneault, F., Cheriet, M.: Blockchain based model for consent management and data transparency assurance. In: 21st International Conference on Software Quality, Reliability and Security Companion, QRS-C 2021, Virtual, Hainan, pp. 1050–1059. IEEE (2021). https://doi.org/10.1109/QRS-C55045.2021.00159

Anish, J.: Data Protection and Privacy Laws & Regulations in India (2024). https://burgeon.co.in: https://burgeon.co.in/blog/data-protection-and-privacy-laws-and-regulations-in-india/. Accessed 14 Jul 2024

Arun, P.: The TRAI Recommendations on Privacy, Security, and Ownership of Data in the Telecom Sector (2018). https://corporate.cyrilamarchandblogs.com: https://corporate.cyrilamarchandblogs.com/2018/07/trai-recommendations-privacy-security-ownership-data-telecom-sector-2018/. Accessed 10 Jul 2024

Azad, M.A., Abdullah, S., Arshad, J., Lallie, H., Ahmed, Y.H.: Verify and trust: a multidimensional survey of zero-trust security in the age of IoT. Internet Things (Netherlands) **27** (2024). https://doi.org/10.1016/j.iot.2024.101227

Barone, C., Mekni, M.: Advancing continuous authentication using smart real-time user activity fingerprinting. In: 3rd IEEE International Conference on Signal, Control and Communication, SCC 2023, Hammamet. IEEE (2023). https://doi.org/10.1109/SCC59637.2023.10527685

Basta, N., Ikram, M., Kaafar, M.A., Walker, A.: Towards a zero-trust micro-segmentation network security strategy: an evaluation framework. In: IEEE/IFIP Network Operations and Management Symposium, NOMS 2022, Budapest. IEEE (2022). https://doi.org/10.1109/NOMS54207.2022.9789888

BBC: Marriott hack hits 500 million Starwood guests (2018). https://www.bbc.com: https://www.bbc.com/news/technology-46401890. Accessed 22 Jul 2024

Bensari, M., Bilami, A.: A new hybrid authentication protocol to secure data communications in mobile networks, vol. 488 (2013). https://doi.org/10.1007/978-3-319-00560-7_24

Bhatt, R., Gupta, M., Sharman, R.: Identity management systems: models, standards, and COTS offerings. In: Identity Theft: Breakthroughs in Research and Practice, pp. 129–155. IGI Global (2016). https://doi.org/10.4018/978-1-5225-0808-3.ch007

Bhatt, S., Santhanam, T.: Keystroke dynamics for biometric authentication-a survey. In: International Conference on Pattern Recognition, Informatics and Mobile Engineering, PRIME 2013, pp. 17–23 (2013). https://doi.org/10.1109/ICPRIME.2013.6496441

Bolognini, L., Bistolfi, C.: Pseudonymization and impacts of Big (personal/anonymous) data processing in the transition from the Directive 95/46/EC to the new EU General Data Protection Regulation. Comput. Law Secur. Rev. **33**(02), 171–181 (2017). https://doi.org/10.1016/j.clsr.2016.11.002

Bumiller, A., Barais, O., Challita, S., Combemale, B., Aillery, N., Le Lan, G.: A context-driven modelling framework for dynamic authentication decisions. In: 48th Euromicro Conference on Software Engineering and Advanced Applications, SEAA 2022, Gran Canaria, pp. 398–405. Institute of Electrical and Electronics Engineers Inc. (2022). https://doi.org/10.1109/SEAA56994.2022.00069

Cabinakova, J., Ostern, N.K., Krönung, J.:. Understanding preprototype user acceptance of centralised and decentralised identity management systems. In: 27th European Conference on Information Systems: Information Systems for a Sharing Society, ECIS 2019. Stockholm and Uppsala: Association for Information Systems (2019). https://www.scopus.com/record/display.uri?eid=2-s2.0-85087109398&origin=scopusAI. Accessed 10 Jul 2024

Carmel, V.V., D., A.: A survey on biometric authentication systems in cloud to combat identity theft. J. Critical Rev. **07**(03), 540–547 (2020). https://doi.org/10.31838/jcr.07.03.97

Carole, C., Emma, G.-H.: Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach (2018). https://www.theguardian.com: https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election. Accessed 22 Jul 2024

Charan Ayineni, K.S., Pranav, P.M., Vasanth, A.V.: Enhancing authentication security: emoji based graphical passwords in universal three-factor authentication systems. In: 7th International Conference on Inventive Computation Technologies, ICICT 2024, pp. 2152–2159. Institute of Electrical and Electronics Engineers Inc. (2024). https://doi.org/10.1109/ICICT60155.2024.10544476

Chatterjee, S., Ghosh, S.K., Chaudhuri, R., Chaudhuri, S.: Adoption of AI-integrated CRM system by Indian industry: from security and privacy perspective. Inf. Comput. Secur. **29**(01), 01–24 (2020). https://doi.org/10.1108/ICS-02-2019-0029

De, S.J., Shukla, R.: Privacy policies of e-governance initiatives: evidence from India. J. Public Affairs **20**(04) (2020). https://doi.org/10.1002/pa.2160

Deborah, S.A., Rajalakshmi, S., Mirnalinee, T.: Secure identity, access, and mobility management in the Internet of Things. In: Secure Communication in Internet of Things: Emerging Technologies, Challenges, and Mitigation, pp. 107–116. CRC Press (2024). https://doi.org/10.1201/9781003477327-9

Dhiman, P., et al.: A review and comparative analysis of relevant approaches of zero trust network model. Sensors **24**(04) (2024). https://doi.org/10.3390/s24041328

Drozd, O., Kirrane, S.: Privacy CURE: consent comprehension made easy. In: 35th IFIP TC 11 International Conference on Information Security and Privacy Protection, SEC 2020. 580 IFIP, Maribor, pp. 124–139. Springer Science and Business Media Deutschland GmbH (2020). https://doi.org/10.1007/978-3-030-58201-2_9

Fan, Y.: Paradigm review of data localization in India and its implications for China. Bull. Chin. Acad. Sci. **38**(08), 1177–1186 (2023). https://doi.org/10.16418/j.issn.1000-3045.20230318001

Freeman, J., Park, S.: Rural realities: digital communication challenges for rural Australian local governments. Transforming Govern. People Process Policy **09**(04), 465–479 (2015). https://doi.org/10.1108/TG-03-2015-0012

Freeman, J., Park, S., Middleton, C.: Technological literacy and interrupted internet access. Inf. Commun. Soc. **23**(13), 1947–1964 (2020). https://doi.org/10.1080/1369118X.2019.1623901

Gupta, S., Jain, S.: An efficient encryption algorithm for sensitive data using numeric and alphanumeric format. J. Comput. Theor. Nanosci. **15**, 2897–2907 (2018). https://doi.org/10.1166/jctn.2018.7561

Hassan, S., Ahmad, R., Katuk, N., Ghazali, N.N., Aripin, J.A., Ali, F.: Staying one step ahead: exploring protection motivation theory to combat cyber-fraud among e-services users. In: 7th Information Systems International Conference, ISICO 2023, vol. 234, pp. 1364–1371. Elsevier B.V., Washington (2024). https://doi.org/10.1016/j.procs.2024.04.011

Ibrahim, D.R., Tamimi, A.A., Abdalla, A.M.: Performance analysis of biometric recognition modalities. In: 8th International Conference on Information Technology, ICIT 2017, Amman, pp. 980–984. Institute of Electrical and Electronics Engineers Inc. (2017). https://doi.org/10.1109/ICITECH.2017.8079977

IGA, K.: Facebook and data privacy in the age of Cambridge analytica. https://jsis.washington.edu: https://jsis.washington.edu/news/facebook-data-privacy-age-cambridge-analytica/. Accessed 22 Jul 2024

Jain, N., Cherukuri, A.K., Kamalov, F.: Revisiting fully homomorphic encryption schemes for privacy-preserving computing. In: Emerging Technologies and Security in Cloud Computing, pp. 276–294. IGI Global (2024). https://doi.org/10.4018/979-8-3693-2081-5.ch012

Jawad, L.A.: Security and privacy in digital healthcare systems: challenges and mitigation strategies. Abhigyan **42**(01), 23–31 (2024). https://doi.org/10.1177/09702385241233073

Julia, C.W.: Uber concealed massive hack that exposed data of 57m users and drivers (2017). https://www.theguardian.com: https://www.theguardian.com/technology/2017/nov/21/uber-data-hack-cyber-attack. Accessed 22 Jul 2024

Kalyan, P.: Economic Survey 2023–2024: India among world's fastest growing 5G networks (2024). https://economictimes.indiatimes.com: https://economictimes.indiatimes.com/industry/telecom/telecom-news/economic-survey-2023-2024-india-among-worlds-fastest-growing-5g-networks/articleshow/111921763.cms. Accessed 25 Jul 2024

Karim, N.A., Kanaker, H., Abdulraheem, W.K., Ghaith, M.A., Alhroob, E., Alali, A.M.: Choosing the right MFA method for online systems: a comparative analysis. Int. J. Data Network Sci. **08**(01), 201–212 (2024). https://doi.org/10.5267/j.ijdns.2023.10.003

Kawase, R., Diana, F., Czeladka, M., Schüler, M., Faust, M.: Internet fraud: the case of account takeover in online marketplace. 30th ACM Conference on Hypertext and Social Media, HT 2019, Hof, pp. 181–190. Association for Computing Machinery, Inc. (2019). https://doi.org/10.1145/3342220.3343651

Kenneth, A., et al.: Phishing attack awareness among college students. In: 20233rd International Conference on Electronic and Electrical Engineering and Intelligent System, ICE3IS 2023, Hybrid, Yogyakarta, pp. 344–348. Institute of Electrical and Electronics Engineers Inc. (2023). https://doi.org/10.1109/ICE3IS59323.2023.10335412

Khan, E., Meraj, S., Khan, M.M.: Security analysis of QKD protocols: simulation comparison. In: 17th International Bhurban Conference on Applied Sciences and Technology, IBCAST 2020, Islamabad, pp. 383–388. IEEE (2020). https://doi.org/10.1109/IBCAST47879.2020.9044522

Kobielus, J.: Big Identity's double-edged sword: wielding it responsibly. IBM Data Management Magazine (12) (2023). https://www.scopus.com/record/display.uri?eid=2-s2.0-84896891808&origin=scopusAI. Accessed 10 Jul 2024

Kulkarni, P., Akhilesh, K.B.: Role of Cyber Security in Public Services. In: Smart Technologies: Scope and Applications, pp. 67–77. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-7139-4_5

Labadie, C., Legner, C.: Personal data management inside and out Integrating data protection requirements in the data life cycle. Enterp. Modell. Inf. Syst. Archit. **15** (2020). https://doi.org/10.18417/emisa.15.9

Li, L.: Data security technology in electronic commerce system development. In: 1st IEEE International Conference on Applied Intelligence and Sustainable Computing, ICAISC 2023, Hybrid, Dharwad: Institute of Electrical and Electronics Engineers Inc. (2023). https://doi.org/10.1109/ICAISC58445.2023.10200851

Ludera, D.T.: Credit card fraud detection by combining synthetic minority oversampling and edited nearest neighbours. In: Future of Information and Communication Conference, FICC 2021. 1364 AISC, pp. 735–743. Virtual, Online: Springer Science and Business Media Deutschland GmbH (2021). https://doi.org/10.1007/978-3-030-73103-8_52

Mandal, A., Amilan, S.: Fathoming fraud: unveiling theories, investigating pathways and combating fraud. J. Finan. Crime (2023). https://doi.org/10.1108/JFC-06-2023-0153

Markkanen, V., Frantti, T.: Secure 3rd party software integration. In: 10th International Conference on Dependable Systems and Their Applications, DSA 2023, Tokyo, pp. 250–259. Institute of Electrical and Electronics Engineers Inc (2023). https://doi.org/10.1109/DSA59317.2023.00038

Maurya, H., Prasad, S.: Data protection laws and a comparative analysis of GDPR and PDPB. In: 1st International Conference on Advances in Computational Science and Engineering, ICACSE 2020, Coimbatore, vol. 2519. American Institute of Physics Inc. (2020). https://doi.org/10.1063/5.0110597

Memon, I.: A secure and efficient communication scheme with authenticated key establishment protocol for road networks. Wirel. Personal Commun. **85**(03), 1167–1191 (2015). https://doi.org/10.1007/s11277-015-2833-0

Milisavljević, S., Mitrović, S., Nešić, L.G., Šimunović, G., Kozak, D., Antić, A.: The level of correlation between cultural values and system of customer relationship management. Tehnicki Vjesnik, 20(06), 1037–1042 (2013). https://www.scopus.com/record/display.uri?eid=2-s2.0-84890935777&origin=scopusAI. Accessed 10 Jul 2024

Naik, N., Grace, P., Jenkins, P.: An attack tree based risk analysis method for investigating attacks and facilitating their mitigations in self-sovereign identity. In: IEEE Symposium Series on Computational Intelligence, SSCI 2021, Orlando. IEEE (2021). https://doi.org/10.1109/SSCI50451.2021.9659929

Nedungadi, P.P., Menon, R., Gutjahr, G., Erickson, L., Raman, R.: Towards an inclusive digital literacy framework for digital India. Education Train. **60**(06), 516–528 (2018). https://doi.org/10.1108/ET-03-2018-0061

Online, B.: Yahoo Data Breach: What Actually Happened? (2020). https://bpbonline.medium.com: https://bpbonline.medium.com/yahoo-data-breach-what-actually-happened-54cf8f3f7c93. Accessed 12 Jul 2024

Pavneet Singh, C.: In new-age data theft, fraudsters steal thumb prints from land registries. https://indianexpress.com: https://indianexpress.com/article/cities/delhi/in-new-age-data-theft-fraudsters-steal-thumb-prints-from-land-registries-7914530/. Accessed 12 Jul 2024

Podugu, S., Rayapureddi, V.K., Gupta, M.: Auditing customer identity and access management. In: Podugu, S., Rayapureddi, V.K., Gupta, M. (eds.) Modernizing Enterprise IT Audit Governance and Management Practices, pp. 181–210. IGI Global (2023). https://doi.org/10.4018/978-1-6684-8766-2.ch007

Pohn, D., Hommel, W.: New directions and challenges within identity and access management. IEEE Commun. Standards Magazine **07**(02), 84–90 (2023). https://doi.org/10.1109/MCOMSTD.0006.2200077

Poritskiy, N., Oliveira, F., Almeida, F.: The benefits and challenges of general data protection regulation for the information technology sector. Digital Policy, Regulation Governance **21**(05), 510–524 (2019). https://doi.org/10.1108/DPRG-05-2019-0039

Pramila, R., Misbahuddin, M., Shukla, S.: A survey on adaptive authentication using machine learning techniques. In: International Conference on Data Science, Computation, and Security, IDSCS 2022. vol. 462, pp. 317–335. Springer, Lavasa (2022). https://doi.org/10.1007/978-981-19-2211-4_28

Purwanto, A., Soewito, B.: Optimization problem of computer network using ppdioo. ICIC Express Lett. **15**(07), 769–777 (2021). https://doi.org/10.24507/icicel.15.07.769

Rasouli, H., Valmohammadi, C.: Proposing a conceptual framework for customer identity and access management: a qualitative approach. Global Knowl. Memory Commun. **69**(1–2), 94–116 (2020). https://doi.org/10.1108/GKMC-02-2019-0014

Raudha, F., Saeedi, M.: Artificial intelligence and machine learning as a tool in preventing and detecting financial fraud: a systematic literature review. J. Adv. Res. Dyn. Control Syst. **11**(11), 904–911 (2019). https://doi.org/10.5373/JARDCS/V11SP11/20193114

Samatha, J., Madhavi, G.: Securesense: enhancing person verification through multimodal biometrics for robust authentication. Scalable Comput. **25**(02), 1040–1055 (2024). https://doi.org/10.12694/scpe.v25i2.2524

Senarath, D., Tharinda, S., Vishvajith, M., Rasnayaka, S., Wickramanayake, S., Meedeniya, D.: BehaveFormer: a framework with spatio-temporal dual attention transformers for IMU-enhanced keystroke dynamics. In: IEEE International Joint Conference on Biometrics, IJCB 2023. Ljubljana: IEEE (2023). https://doi.org/10.1109/IJCB57857.2023.10448997

Shaik, A., Borgaonkar, R., Park, S., Seifert, J.-P.: New vulnerabilities in 4G and 5G cellular access network protocols: exposing device capabilities. In: 12th Conference on Security and Privacy in Wireless and Mobile Networks, WiSec 2019, pp. 221–232. Miami: Association for Computing Machinery, Inc. (2019). https://doi.org/10.1145/3317549.3319728

Shehu, A.-S., Pinto, A., Correia, M.E.: Providing secured access delegation in identity management systems. In: 17th International Conference on Security and Cryptography, SECRYPT 2020 - Part of the 17th International Joint Conference on e-Business and Telecommunications, ICETE 2020. 03, pp. 638–644 (2020). Online: SciTePress. https://doi.org/10.5220/0009892206380644

Simha, R., Krantz, H.K., R., Prabhu, A., Joshi, P.: Beyond passwords: a multi-factor authentication approach for robust digital security. Internet Technol. Lett. (2024). https://doi.org/10.1002/itl2.555

Simonds, V.W., Goins, R.T., Krantz, E.M., Garroutte, E.M.: Cultural identity and patient trust among older American Indians. J. General Internal Med. **29**(03), 500–506 (2014). https://doi.org/10.1007/s11606-013-2578-y

Singh, S., Sirohi, N.J.: Internet banking services as tool of CRM: a study of customer satisfaction in the National Capital Region, Delhi. In. J. Electron. Custom. Relat. Manag. **08**(01–03), 118–123 (2014). https://doi.org/10.1504/IJECRM.2014.066888

Soomro, Z.A., Ahmed, J., Shah, M.H., Khoumbati, K.: Investigating identity fraud management practices in e-tail sector: a systematic review. J. Enterp. Inf. Manag. **32**(02), 301–324 (2019). https://doi.org/10.1108/JEIM-06-2018-0110

Soviany, C.: The benefits of using artificial intelligence in payment fraud detection: a case study. J. Payments Strategy Syst. **12**(02), 102–110 (2018). https://www.scopus.com/record/display.uri?eid=2-s2.0-85059097189&origin=scopusAI. Accessed 20 Jul 2024

Steenbergen, O., Meesters, K.: Towards transparentizing personal data processing, winning public trust. In: 42nd International Conference on Information Systems: Building Sustainability and Resilience with IS: A Call for Action, ICIS 2021 TREOs, Austin. Association for Information Systems (2021). Retrieved Jul 03, 2024, from https://www.scopus.com/record/display.uri?eid=2-s2.0-85175422041&origin=scopusAI

Sujith, S.: Cybersecurity Spotlight: Analyzing the Marriott Data Breach and Its Impact (2024). https://www.linkedin.com: https://www.linkedin.com/pulse/cybersecurity-spotlight-analyzing-marriott-data-breach-selvaraj-x606e/. Accessed Jul 13, 2024

Szatmáry, S.: Quantum computers—security threats and solutions. In: Advanced Sciences and Technologies for Security Applications, pp. 431–441. Springer (2024). https://doi.org/10.1007/978-3-031-47990-8_38

Taher, K. A., Nahar, T., Hossain, S.A.: Enhanced cryptocurrency security by time-based token multi-factor authentication algorithm. In: 1st International Conference on Robotics, Electrical and Signal Processing Techniques, ICREST 2019, Dhaka, pp. 308–312. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICREST.2019.8644084

Tamanna, Kamboj, S., Singh, L., Kaur, T.: Automated fraud detection in financial transactions using machine learning: an ensemble perspective. In: 2nd International Conference on Artificial Intelligence and Machine Learning Applications, AIMLA 2024, Namakkal. Institute of

Electrical and Electronics Engineers Inc. (2024). https://doi.org/10.1109/AIMLA59606.2024.10531422

Thorve, A., Shirole, M., Jain, P., Santhumayor, C., Sarode, S.: Decentralized identity management using blockchain. In: 4th International Conference on Advances in Computing, Communication Control and Networking, ICAC3N 2022, Greater Noida, pp. 1985–1991. IEEE (2022). https://doi.org/10.1109/ICAC3N56670.2022.10074477

Tokas, S., Owe, O.: A formal framework for consent management. In: 40th IFIP WG 6.1 International Conference on Formal Techniques for Distributed Objects, Components, and Systems, FORTE 2020, held as part of the 15th International Federated Conference on Distributed Computing Techniques, DisCoTec 2020. 12136 LNCS, pp. 169–186. Springer, Valletta (2020). https://doi.org/10.1007/978-3-030-50086-3_10

Tripathi, R., Tripathi, S.: Frauds and cyber security issues in the finance sector. In: Transforming the Financial Landscape With ICTs, pp. 165–189 (2024). IGI Global. https://doi.org/10.4018/979-8-3693-1503-3.ch008

Tzafilkou, K., Protogeros, N., Yakinthos, C.: End-user development of CRM systems: towards a behavioural end-user profiling based on gender and expertise. Int. J. Electron. Bus. **12**(03), 302–318 (2015). https://doi.org/10.1504/IJEB.2015.071390

Vasishtan, P.: The Passport to Regulate Foreign Jurisdiction: The Personal Data Protection Bill, 2019 on its Extraterritorial Application. Brawijaya Law J. **09**(01), 47–58 (2022). https://doi.org/10.21776/ub.blj.2022.009.01.04

Vendome, C., Solano, D., Linan, S., Linares-Vasquez, M.: Can everyone use my app? An empirical study on accessibility in android apps. In: 2019 IEEE International Conference on Software Maintenance and Evolution, ICSME 2019, pp. 41–52. Institute of Electrical and Electronics Engineers Inc., Cleveland (2019). https://doi.org/10.1109/ICSME.2019.00014

Viguri Cordero, J.A.: The use of certification mechanisms as an efficient guarantee of personal data protection. Revista Catalana de Dret Public **2021**(62), 160–176 (2021). https://doi.org/10.2436/rcdp.i62.2021.3571

Vivek, B., Nandhan, S.H., Zean, J.R., D., L., Dhanwanth, B.: Applying machine learning to the detection of credit card fraud. Int. J. Intell. Syst. Appl. Eng. **11**(03), 643–652 (2023). https://www.scopus.com/record/display.uri?eid=2-s2.0-85174806778&origin=scopusAI. Accessed Jul 10, 2024

Wiercioch, A., Teufel, S., Teufel, B.: The authentication dilemma. J. Commun. **13**(08), 443–449 (2018). https://doi.org/10.12720/jcm.13.8.443-449

Xu, B., Wang, Y., Liao, X., Wang, K.: Efficient fraud detection using deep boosting decision trees. Dec. Support Syst. **175** (2023). https://doi.org/10.1016/j.dss.2023.114037

Yadav, D., Mishra, M., Prakash, S.: Mobile forensics challenges and admissibility of electronic evidences in India. In: 5th International Conference on Computational Intelligence and Communication Networks, CICN 2013, pp. 237–242. Sergeant Infotech Pvt. Ltd. (2013). https://doi.org/10.1109/CICN.2013.57

Yalamuri, G., Honnavalli, P., Eswaran, S.: A review of the present cryptographic arsenal to deal with post-quantum threats. In: 4th International Conference on Innovative Data Communication Technologies and Application, ICIDCA 2022. 215, pp. 834–845. Elsevier B.V., Coimbatore (2022). https://doi.org/10.1016/j.procs.2022.12.086

Yan, Y., Liu, S., Jin, Y., Qian, Z., Zhang, S., Lu, S.: Risk minimization against transmission failures of federated learning in mobile edge networks. IEEE Access **08**, 98205–98217 (2020). https://doi.org/10.1109/ACCESS.2020.2996307

MERN Stack Website for Government Officials

Rohit A. Patil[1(✉)], Kalpana S. Pawase[1], Jyoti Sarwade[2], Kaustubh Chaudhari[2], and Tushar Suryawanshi[2]

[1] EnTC, MIT Academy of Engineering, Pune 412105, Maharashtra, India
{rohit.patil,kalpana.pawase}@mitaoe.ac.in
[2] Electronics, MIT Academy of Engineering, Pune 412105, Maharashtra, India
{jyoti.sarwade,kaustubh.chaudhari,
tushar.suryawanshi}@mitaoe.ac.in

Abstract. In today's life, with the hectic pace, it is very important to have effective communication in due time, particularly in administration where decisions are made to promote the well-being of thousands of people. To this end, however, many government officials are still using very old communication tools that cannot set them on the same page, and a decision cannot be taken in due time. Commulink is likely to fill this gap, constituting a revolutionary communications platform to change the way government staff could work as well as communicate with one another.

Commulink facilitates communication in different government sectors in such a way that interaction is fast, safe, and in real-time. Through this platform, various sector employees can easily interact with each other without the usual hassle, ensuring smooth, timely decision-making processes. With a suite of robust features-from instant messaging, safe file sharing, and video conferencing to the integrated task management Commulink streamlines communication but makes things much more efficient at the operational level.

One feature of Commulink is inductivity. It really focuses on accessibility, where everyone, regardless of their disability, can participate in a communication process. Thus, Commulink is considered a versatile solution for diverse teams because of equal engagement for all employees.

Additionally, Commulink emphasizes security concerns; it is, after all, very sensitive information that involves government departments. Information, by end-to-end encryption and more advanced security protocols, is safeguarded against cyber threats from fraudsters, data thieves, and hackers-and a sense of confidence is instilled in these agencies for future communication. In today's life, with the hectic pace, it is very important to have effective communication in due time, particularly in administration where decisions are made to promote the well-being of thousands of people. To this end, however, many government officials are still using very old communication tools that cannot set them on the same page, and a decision cannot be taken in due time. Commulink is likely to fill this gap, constituting a revolutionary communications platform to change the way government staff could work as well as communicate with one another.

K. S. Pawase, J. Sarwade, K. Chaudhari and T. Suryawanshi—Contributing authors. These authors contributed equally to this work.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 302–308, 2025.
https://doi.org/10.1007/978-3-031-86305-9_20

Commulink facilitates communication in different government sectors in such a way that interaction is fast, safe, and in real-time. Through this platform, various sector employees can easily interact with each other without the usual hassle, ensuring smooth, timely decision-making processes. With a suite of robust features-from instant messaging, safe file sharing, and video conferencing to the integrated task management-Commulink streamlines communication but makes things much more efficient at the operational level.

One feature of Commulink is inductivity. It really focuses on accessibility, where everyone, regardless of their disability, can participate in a communication process. Thus, Commulink is considered a versatile solution for diverse teams because of equal engagement for all employees.

Commulink emphasizes security concerns; it is, after all, very sensitive information that involves government departments. Information, by end-to-end encryption and more advanced security protocols, is safeguarded against cyber threats-from fraudsters, data thieves, and hackers-and a sense of confidence is instilled in these agencies for future communication.

1 Introduction

Effective communication is essential in any government organization. Traditional methods, though, in most cases, cannot provide a secure, real-time, and collaborative platform for government officials. We present Commulink, a comprehensive communication platform implemented using the MERN stack. This paper discusses the design, implementation, and evaluation of Commulink with respect to how it will improve government operations by enhancing collaboration and information sharing.

Commulink is looking to change the nature of interaction and information dissemination among government officials through a suite of features that includes real-time messaging, file sharing, audio and video conferencing—all bound together with robust security provisions.

2 Literature Survey

Investigations of government communication platforms bring forth the acute need for effective, secure, and transparent channels in an official dissemination of information. On the other hand, conventional systems have certain limitations; thus, calling for innovative solutions. MERN stack happens to be among the technologies through which reliable and scalable communication networks can be achieved, together with MongoDB, Express.js, React, and Node.js. Literature pointed out that it is also of high security protocols, especially in dealing with sensitive information. Accessibility features reach out to the users who have disabilities. From literature, advanced communication platforms have been perceived as holding the key to effective decision-making and operational efficiency within the government setup.

2.1 Social Media Web App Using MERN

This research paper deals with the application of the MERN stack in dynamic web development. This paper presents how social media has been an effective platform for communication, wherein several pros and cons associated with it are discussed.

2.2 MERN Stack Web App for College

Details the development of a MERN-based online application for college administration, covering all associated aspects of costs, attendance, academics, resources, and alumni engagement.

2.3 Real-Time Chat Website

A real-time chat application developed using the MERN stack, focusing on high- level data security and user authentication, messaging, and setup of chat rooms and accounts.

2.4 MERN Stack Chat Website

The chat application will be a modern day communication-driven chat application implemented using the MERN technologies that provide all security features, real-time interaction, and user-friendly interfaces.

2.5 Performance Optimization Using MERN Stack on Web Application

This paper focuses on the development of a MERN e-commerce application and how each technology has its role to play in smoothening online shopping and business performance.

3 Project Planning

3.1 Project Objectives

Design a secure real-time communication platform for government officials. Mix messaging, file sharing, and au dio/video calls with member management. Provide access through hands-free and blind modes for very different dis abilities. Develop and be flexible in MERN Stack.

3.2 Project Scope

Front-end with React, Redux, TailwindCSS, ShadcnUI; back-end with Node.js, Express.js, Socket.io, Prisma ORM; MySQL database management with Planetscale; integration with third-party services, among them UploadThing and Clerk; real-time functionality realized by means of Socket.io; accessible UI development for persons with disabilities; and unit test and QA.

3.3 Project Timeline

Phase 1 was Requirement gathering, system design, and prototyping. Phase 2 was Frontend development with API integration. Phase 3 was Integration of back-end development with database setup. Phase 4 was Real-time features implementation and testing. Phase 5 was Apply accessibility features and test. Phase 6 was Deployment and User Acceptance Testing.

4 Methodology

4.1 System Development Methodology

Agile, more precisely Scrum, underpinned the development process of Commulink. This helped to provide a really dynamic and adaptive development environment. Its iterative approach allowed for the rapid prototyping of features in view of their continuous evaluation and refinement according to user feedback and evolving requirements. Breaking down the project into manageable sprints allowed the development team to ensure an efficient workflow for the timely delivery of key functionalities.

It is in this aspect that Scrum places great importance on collaboration and transparency, which was critical in binding the development environment together. Ensuring open communication and knowledge sharing among all members of the team was done through daily stand-up meetings, sprint planning sessions, and sprint retrospectives. Such a collaborative approach makes problem-solving and decision-making easier and aligns development efforts toward common goals.

4.2 Technology Stack

To build a solid, scalable, and maintainable platform, Com mulink used the MERN stack as its core technology. This stack, enriched by MongoDB, Express.js, React, and Node.js, gave developers a full-fledged toolkit at their disposal for building both frontend and backend components. In this case, it was the flexible schema and high performance of MongoDB that were instrumental in managing such diverse data requirements on the platform. Afterward, Express.js will help smooth out the development of the API and server-side logic. React, due to its component architecture, was helpful in creating a dynamic and user-friendly interface for improving the overall user experience. Node.js served as the runtime environment by providing an efficient and scalable execution platform.

Other technologies complemented the core functionalities. Socket.io facilitated the set-up of real-time channels used for messaging, notifications, and collaborative editing. Prisma ORM smoothened interactions with the database, hence improving efficiency in development and reducing probable errors. Tailwind CSS and Shadcn UI were used to create a nice responsive user interface in accordance with modern design principles.

4.3 System Architecture

Commulink's architecture is based on the client-server model, wherein the frontend and backend components communicate to provide a rich user experience. On the other hand, the frontend—driven by React—is in charge of user interactions, interface rendering, and local state management. The backend, powered by Node.js with Express.js, processes requests, manages data, and runs business logic. MongoDB stands as the main database used in storing user information, messages, and other relevant data.

The role of Socket.io in enabling real-time communication between clients and the server is very instrumental. Socket.io establishes persistent connections through which efficient data exchange is enabled, hence users' updating and notifying with much ease. That is very critical for real-time messaging or collaborative editing features or even a simple 'presence indicator'.

4.4 Development Process

The development of Commulink followed a guided process through a number of very important phases:

Requirement Gathering and Analysis: For this to work, proper tar- get user needs understanding and an in-depth understanding of the requirements in government communication were necessary. Then, a detailed stakeholder inter- view, workshops, and document analysis helped to define the comprehensive set of requirements.

System Design: Based on the requirements gathered, the entire system architecture had to be worked out; defining system components, modules, and their interaction. It included database schema creation, user interface wireframing, and API specification definition.

Development and Implementation: The development team adopted an Agile approach wherein the platform was developed iteratively. At the same time, frontend and backend development were done in parallel with close collaboration to make sure seamless integration is delivered. Testing and Quality Assurance Testing at each level of the development process was vigorously conducted to ensure the quality and reliability of the platform and its security. Unit, integration, and system testing were done for defect identification and resolutions.

Deployment: After testing and validation, the platform was deployed to a secure cloud infrastructure to ensure scalability and accessibility. Maintenance and Support: The platform has been under constant monitoring and has been updated with technical support to keep it running and to address user needs and evolving requirements.

5 Output Images

See the Figs. 1, 2, 3 and 4.

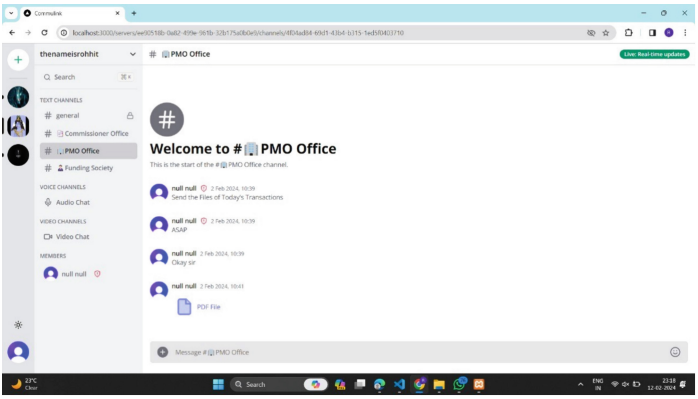


Fig. 1. Chatting Interface

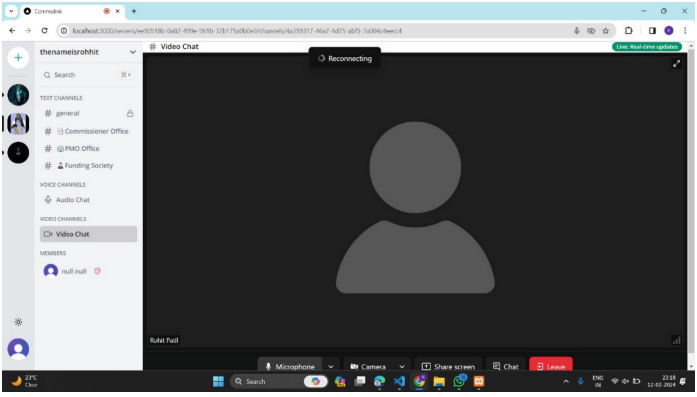


Fig. 2. Video conferencing Interface

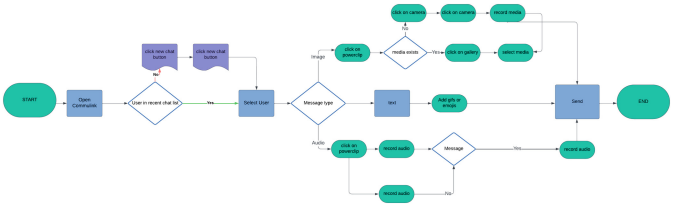


Fig. 3. Working flow chart

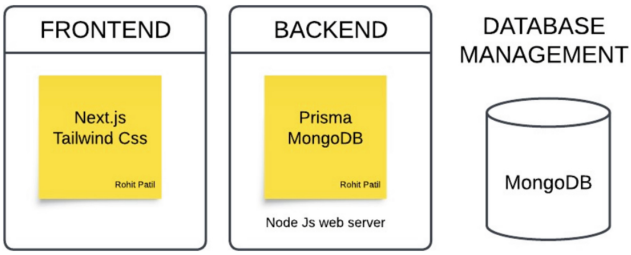


Fig. 4. Technologies used

Acknowledgment. In particular, thanks go to those people who have supported and guided us over the semester. The valuable assistance and encouragement that they extended when we were in a fix really came handy to pursue this venture successfully.

We, therefore, the first, foremost, and so very sincerely place on record our thanks to Dr. Kalpana Pawase, our project adviser, for her unwavering support, insightful advice, and expert guidance. Her wisdom, patience, and dedication helped us breathe through the challenges and complexities of our project. Her periodic feedback and suggestions have not only enhanced the quality of our work but also motivated us to stretch our horizons and reach for our best.

To name a few, the school dean, Dr. Dipti Sakhare, is to whom one must look for enumeration of gratitude. This person has been a continuous guiding and inspiring force to our whole journey in life. In fact, it would be her marvelous leadership vision that has fostered a productive environment for us here to develop as individuals and professionals. Her trust has made what was always impossible for us possible.

We would love to conclude by thanking all the faculty members and staff for the knowledge, experience, and professional guidance freely donated to us. Their assistance-be it direct collaboration, constructive feedback, or technical aid-made this initiative possible. The dedicated involvement of those persons made the project a success; otherwise, this could not have been achieved.

References

1. Kulkarni, M.V.: Social Media Web Application using MERN. Int. Res. J. Eng. Technol. (IRJET), 09(02) (2022). J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73
2. Nikose, A., Dosani, S., Pardhi, S., Nikode, D., Jais, A.: Real-time chat application. Int. J. Sci. Res. Sci. Eng. Technol. **10**(2), 496–501 (2023). https://doi.org/10.32628/IJSRSET2310267. K. Elissa, "Title of paper if known," unpublished
3. Bawane', M., V.J.R.N. P. S., Gawande. I.: A review on technologies used in MERN stack. Int. J. Res. Appl. Sci. Eng. Technol. (URASET) **10** (2022)
4. Jain, P., Manjunath, C.R.: Comparative analysis of MEAN stack and MERN stack. In: Proceedings of the Journal of Emerging Technologies and Innovative Research (JETIR), May 2021, vol. 8, no. 5 (2021)
5. Tran, H.: Developing a social platform based on MERN stack. In: 2021 Metropolia University of Applied Sciences, Bachelor of Engineering Information Technology (2021)
6. Mehra, M., Kumar, M., Maurya, A., Sharma, C., Shanu: MERN stack web development **25** (2021)

Brain Tumor Diagnosis using Deep Learning: A Systematic Review and Meta-analysis of MRI Image-Based Studies

Patel Kruti Dineshbhai(✉) and Himanshu Maniar

Bhagwan Mahavir University, Surat, Gujarat, India
krutipatel3664@gmail.com, himanshu.maniar@bmusurat.ac.in

Abstract. Traditional methods for the categorization of brain tumors identification and mostly depend on radiologists' time-consuming manual interpretation and subjective. Detecting a brain tumor in its early stages can be quite very difficult for doctors. Traditional methods for the categorization of brain tumors identification and mostly depend on radiologists' time-consuming manual interpretation and subjective. This process can be subjective, complex and laborious. Recent developments in deep learning have greatly improved the automation and accuracy of medical image processing, including the categorization and identification of brain tumors from MRI scans. The study emphasizes how crucial it is to solve MRI IMAGE dataset: A variety of datasets, including medical datasets, are hosted by Kaggle, a platform for data science competitions (Kaggle offers datasets related to brain tumors at (https://www.kaggle.com/datasets) in MRI image constraints, optimize segmentation algorithms, and incorporate cutting-edge methods like data. MRI pictures may be more susceptible to noise and other environmental disruptions. Doctors find it more challenging to diagnose the tumor and its sources as a result. In order to overcome this, we created a method that uses grayscale image conversion to identify brain tumors from images. To remove noise and other distracting elements from the image, we employ filters. The system's processing will include a preprocessing step for the selected image. To find brain cancers in the MRI pictures, multiple algorithms are applied at the same time. However, the edges of the image may not be crisp and distinct in the early stages of a brain tumor. As a result, to identify the edges in the pictures, we are using image segmentation. We have developed a number of filtering algorithms and segmentation procedures to extract information from photos. Precision can be increased overall with this strategy.

Keywords: Brain Tumor · Magnetic Resonance Image (MRI) · Classification · Segmentation

1 Introduction

The brain is part of the human central nervous system. The majority of conventional methods for identifying and classifying brain cancers depend on radiologists' laborious, subjective manual image interpretation. Recent advancements in the deep learning algorithm class known as convolutional neural networks (CNNs) have made it feasible to

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 309–319, 2025.
https://doi.org/10.1007/978-3-031-86305-9_21

automatically assess diagnostic imaging, including the detection and categorization of MRI brain tumors. The standard method of identifying cancers in brain scans involves a human inspection procedure that is unreliable and can yield a range of results depending on the situation [1].

For a trustworthy and strong analysis that is independent of different circumstances, automating the brain imaging tumor detection approach is essential [2]. It is essential to have a solid understanding of the human neurological system before applying image processing algorithms to identify and categorize malignancies from brain MRI data. The initial category covers tumors arising in the central nervous system (CNS), whereas the subsequent category pertains to those in the peripheral nervous system (PNS). The World Health Organization (WHO) states that the diagnosis and classification of the magnetic resonance imaging is dependent upon the cells of origin, the area of the brain they affect, and other factors.

Are based on the detection, categorization, and size of brain tumors derived that pose a hazard, and these tumors have been detected on magnetic resonance imaging (MRI) brain images. This is because MRI offers better contrast for soft tissues in the brain compared to CT scans. Recent advances in medical imaging focus on real-time tumor detection using improved algorithms, with ongoing research concentrating on brain tumors in MRI and CT images. The key to developing an effective clinical procedure is separation, or scalability Deep learning significantly enhances the diagnosis and detection of brain cancer [3]. Moreover, the Probabilistic Neural Network (PNN) is the most effective classification method for brain cancer detection among deep learning techniques [4]. Subsequently The use of PNN in brain cancer detection and diagnosis complements tumor discovery and extraction, relying on algorithms for classification [5]. Artificial intelligence methods such as pattern recognition, fuzzy logic, and machine learning become immensely valuable when coupled. The processing of digital images is one such method. Developing techniques for an automated system to segment brain MRI images is the aim of this effort [6].

The structure of the paper is as follows: reviews related work & covers the strategies for preprocessing and improvement than describes various segmentation strategies and clustering approaches are finally introduced.

2 Related Work

The brain is the most important structural component of brain tumors, which pose a serious worldwide health risk due to their variety of symptoms and difficult-to-treat nature. It is sometimes referred to as the body's center. Researchers in all fields of study, including the medical sciences, employ the technique of image processing, which was created by computer and information technology scientists. The type of pathology, tumor size, and stage all impact treatment options. Experts use computer-aided diagnostics (CAD) and other automated methods to assess the nervous system, as MRI provides higher contrast images for brain soft tissues compared to CT.

A technique utilizing Artificial Neural Networks (ANN) includes steps such image preprocessing, segmentation, feature extraction, SVM classification, and tumor stage classification was presented by Rasel Ahmmed, Anirban Sen Swkshar, Md. Foisal Hossain, and Md. Abdur Rafiq [7]. Important image processing method for brain MRI image

segmentation include SVM and FCM techniques. Swapnil R. Telrandhe et al. [8] proposed a tumor identification approach, emphasizing that segmentation divides an image into distinct regions or objects. This procedure is necessary for accurately classifying the visual content and effectively separating the subject from the background. They conducted experiments to compare various methods, concentrating on the potency of popular edge detection techniques for image segmentation. According to preliminary findings, our strategy has produced successful segmentation results. This approach also considerably lowers the computational requirements.

Brain MR tumors and CT scans are the two areas of research that are now receiving the greatest attention. The latest developments in medical imaging have focused on using more accurate algorithms to diagnose tumors in real time. MRI has been utilized in medicine to distinguish intricate details of the interior architecture of the body. A deep learning module is necessary for the diagnosis of brain malignancies. Brain magnetic resonance imaging (MRI) is a crucial initial step in the identification of brain cancers. An aberrant proliferation of brain cells is represented by a brain tumor. Systems for computer-aided detection (CAD) have come to light as a possible answer to this problem. CAD systems are a useful technique to improve the efficacy and high accuracy of tumor detection in brain MR images. It's imperative to realize, though, that CAD systems are intended to supplement medical professionals?not to replace their expertise and judgment, Important processes such as segmentation, pre-processing images, feature-based classification, and feature extraction are carried out by these models. Deep Learning (DL) methods, particularly Convolutional Neural Networks (CNNs), have gained popularity in the field of medical picture analysis. They outperform more conventional CAD (Computer Aided Design) systems for tasks such as cancer diagnosis (including breast and lung cancer) and tumor classification. More and more people are interested in utilizing state-of-the-art technologies to support early diagnosis and treatment of this potentially lethal illness. The enhanced precision of CNN models and DL algorithms in automating the identification of brain tumors from MRI images is indicative of this (Tables 1 and 2).

3 Gap in Existing Research

Limitations, notably the computational demands associated with training deep neural networks, specially concerning larger datasets. Despite this, they offer insights into mitigating such challenges, such as improving GPU infrastructure to expedite training processes.

Optimization of Segmentation methods: Despite the exploration of a number of segmentation methods, such as fuzzy C-means and k-means clustering, additional optimization and customization are required to address the unique difficulties associated with brain tumor segmentation. Better segmentation accuracy and efficiency could result from research concentrating on improving already-existing algorithms or creating innovative strategies catered to the intricacies of brain tumor imaging. to improved segmentation accuracy and efficiency. The drawback of this study, it is implemented on small dataset methods for processing MRI pictures. The widely recognized statistical method of Principle Component Analysis (PCA), which is used in data analysis and classification, is one way to increase performance. A tiny dataset or uncommon tumor forms may be

Table 1. Literature Review

Reference	Dataset	Models	Performance	Limitations
[9], 2022	10,000 MR images	Deep educational model (proposed), VGG16, ResNet50, MobileNet, Inception V3	Deep Educational model: accuracy 98%	Need to apply Image Augmentation methods
,2022	3394 MR images	Deep convolutional neural network (DCNN) (proposed), VGG16, VGG19, CNN-SVM	DCNN: accuracy 97.72%	Should consider more datasets and various types of images
, 2021	3064, CT brain scan images	CNN	CLM model: Accuracy 96%, Precision and recall 95%	Should consider more datasets to obtain more accurate and promising results
, 2021	2556 brain tumor images	Hybrid ensemble classifier (KNN-RF-DT) (proposed), naive Bayes. Random forest, neural network, KNN, and decision tree	Hybrid Ensemble Classifier (KNN-RF DT): accuracy 97.305%, precision 97.73%, specificity 97.60%, sensitivity 97.04%, Youden index94.71%, F1 score 97.41%	Should consider more datasets to obtain more accurate and promising results
[13], 2022	3260 MRI brain images	Dense EfficientNet (proposed), ResNet 50, MobileNet, MobileNetV2	Dense EfficientNet: Accuracy 98.78%, F1-score 98%	Higher number of parameters and evaluation time
[14], 2022	BRATS 2015 dataset, 2000 MR images	Residual network	Residual network: accuracy of 97.05%	Lack of performance metrics evaluation
[15], 2020 BRATS 2017,	3D image dataset	Modified two-step dragonfly method (proposed), random forest, SVM, fuzzy C-means	Accuracy: 98.20%, recall: 95.13%, and precision: 93.21%	Data processing and over-fitting
[16], 2019	BRATS 2013 MRI Tumor dataset	Hybrid CNN models	Dice score of 86%, sensitivity of 86%, & specificity of 91%	Only one tumor per slice was evaluated
[17], 2019	227 MR images	KNN classifier (proposed), RF, NB, and RBF	Accuracy of 95.6% and AUC of 99%	Lack of MR images

too difficult for current algorithms to handle because of imbalanced datasets or inadequate representation during training Diffusion Tensor imaging (DTI), positron emission tomography (PET), functional MRI (MRI), and other modalities may provide useful information that algorithms fail to consider when focusing only on MRI pictures. Better Segmentation Algorithms: Although there are a lot of segmentation algorithms available, more accurate and healthful techniques are still required, especially when it comes

Table 2. Comparison of Classification Accuracy Obtained by Various Methods

Author	De-noising Method	Segmentation Technique Used	Accuracy in %	
M. Madheswaran et N. al. [19]	ORNRAD filter	Support Vector Machine	SVM-Linear	95.34
			SVM Polynomial	96.51
Eltaher Mohamed Hussein et al. [20]	Adaptive filter & Enhancement filter	Artificial Neural Networks	Recurrent Network (RNN) Design	76.47
			Elman Network Design	88.24
Eman Mohamed Abdel-Maksoud et al. [21]	Median filter	Hybrid clustering technique	K-means	80
			Mean Shift	85.7
			KIFCM	90.5
B. Sathees et al. [22]	Gaussian Filter	Image Mining Techniques	Intensity histogram	89
			GLCM	95.25
Harneet Kaur et al. [23]	Median filter	Object Based Segmentation	Neural	78
			Fuzzy neuron	94
Arashdeep Kaur [24]	High pass Filter	Otsu?s method Thresholding method K-means method Fuzzy c-means method	K-means	89.01
			Otsu?s Method	89.88
			Fuzzy C-means	90.57

to identifying malignancies in brain imaging. Complex problems may be too much for current algorithms to handle.

Early Detection, noisy images, tumor shapes, or variability in image quality. Research focusing on developing more advanced segmentation techniques, Deep learning algorithms can assist researchers in analyzing large Datasets of MR images, leading to a deeper understanding of brain tumors' characteristics, such as deep learning-based approaches, could significantly improve detection accuracy. Research in this area could explore methods for fusing multi-modal images to improve tumor detection sensitivity and specificity.

4 Challenges of the Proposed Research

- Many systems MRI Images are confined only to segmentation.
- They suffer from certain technology issues: - These include problems with processing picture uncertainties that cannot be resolved in any other way, as well as many kinds of incomplete, noisy, imprecise, incompletely dependable, fragmentary, ambiguous, overloaded, contradicting, and insufficient information.

- Some of the systems work only on cancerous image.
- Some of the systems Brain tumor image Staging is not done.
- Less accuracy.
- Area of the tumor isn?t computed.

5 Objective

Deep learning-based algorithm capable of accurately detecting and classification brain tumors from Magnetic resonance (MR) images. Explore and evaluate various deep learning architectures and techniques for image segmentation and classification to enhance the high accuracy and efficiency of brain tumour detection in early stage. Investigate the potential of advanced deep learning models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), or their variants, in detecting subtle and complex features indicative of brain tumors. Enhance the robustness and generalization capabilities of the proposed algorithm by Incorporation techniques such as data augmentation, transfer learning, and ensemble learning. Conduct extensive experimentation and validation on diverse datasets comprising MR images from different modalities and acquisition protocols to assess the algorithm's performance across various scenarios. Compare the performance of the proposed deep learning approach with existing methods for brain tumor detection, including traditional machine learning algorithms and manual segmentation by radiologists. Investigate the interpretability of the deep learning model's predictions to understand the features and patterns driving tumor detection, facilitating clinical decision-making and trust in automated systems. Explore the potential for real-time or near-real-time application of the developed algorithm in clinical settings, considering factors such as computational efficiency and resource requirements. Collaborate with medical professionals and experts in radiology to ensure the clinical relevance, accuracy, and usability of the proposed brain tumor detection system. The dissemination of research findings can be achieved by publishing them in peer-reviewed journals, presenting them at conferences, or storing them in open-access repositories. This will help to enhance medical imaging technology and encourage more research in this area.

6 Proposed System Architecture

See (Fig. 1)

6.1 Data Set

The data collection method involved gathering MRI images of human brain scans from two main sources: the BraTS2013 dataset and the online dataset.

- Research Institutions and Hospitals: Some research institutions and hospitals may Provide access to medical datasets, including those related to brain tumors.

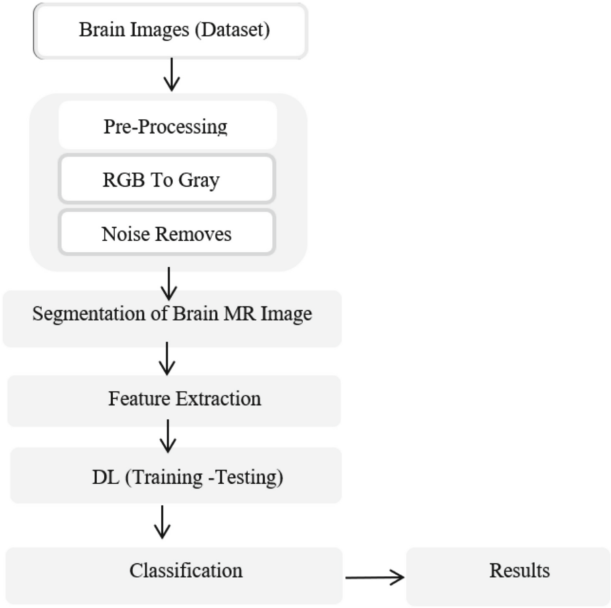


Fig. 1. Brain tumor detection Architecture

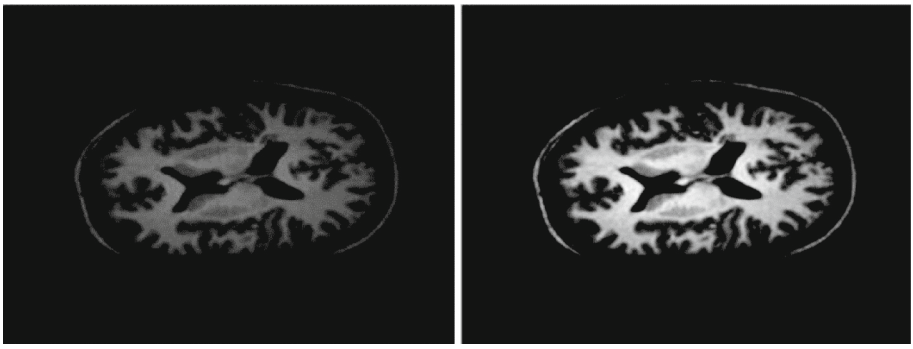
- Kaggle dataset: A variety of datasets, including medical datasets, are hosted by Kaggle, a platform for data science competitions (Kaggle offers datasets related to brain tumors at https://www.kaggle.com/datasets)in MRI image constraints, optimize segmentation algorithms, and incorporate cutting-edge methods like data. Datasets from a variety of fields, including medicine, are hosted by Kaggle, a platform for data science competitions. Kaggle offers datasets related to brain tumors at https://www.kaggle.com/datasets.

6.2 Preprocessing

Techniques for preprocessing and enhancing magnetic resonance imaging (MRIs) are essential to improve the identification of suspicious areas. This section presents the gradient-based image enhancement method for brain magnetic resonance imaging. It uses the first derivative and local statistics data. The first step of the process is removing film artifacts from the MRI, like labels and X-ray marks, using a tracking algorithm. Second, weighted median filtering suppresses high-frequency components and yields higher resolution images than standard median filters, adaptive filters, and spatial filters. The suggested approach is assessed using two metrics: First peak signal-to-noise ratio (PSNR) and second average signal-to-noise ratio (ASNR).

6.3 Segmentation Methods

An important and fundamental stage in image analysis is the use of image segmentation algorithms, which make it possible to extract certain sections from an image. For many



Original Image Enhanced Image

Fig. 2. Pre-processing and Enhancement [25]

medical imaging applications, including computer-assisted surgery, treatment planning, and patient diagnostics, automatic segmentation is crucial. The three primary methods for segmentation are 1.the region-based approach, 2. edge-based approach, and 3. border approach (thresholding). To aid in analysis, segmentation entails breaking an image up into more manageable, smaller pieces. The division of an image into several parts is referred to in this endeavor as segmentation. But because the images are complicated and not usually clearly split, as in MRI scans, or consistently continuous, like in X-ray films, segmentation problems frequently result from this. Each segment in a 2D image is called a pixel; in a 3D image, the term voxel is used instead.

6.4 Feature Extraction

In order to facilitate further analysis, this method entails identifying and presenting particular elements of importance in an image. In many computers vision and imaging applications, it is an essential step. Tumors are classified during feature extraction by taking into account factors like size, shape, composition, and location within the image. The goal of this stage is to extract important features from the input image so that the location of the tumor may be determined through analysis. The MRI image's output prior to the project's feature extraction stage is shown in Fig. 2 above.

7 Possible Outcome

Early Detection: Deep learning algorithms can detect brain tumors before symptoms show up by analyzing MRI data. This allows for earlier intervention and better patient outcomes. Early detection allows for timely intervention and treatment, which frequently improves patient outcomes. Accuracy: Deep learning models are quite good at correctly identifying and classifying different types of brain cancers after being trained on large datasets. This precision lowers the risk of a false positive and makes treatment planning more effective. Efficiency: The time radiologists spend manually analyzing MR images is greatly decreased by deep learning in automated tumor diagnosis, which may result in quicker diagnoses and earlier treatment initiation.

8 Impact of Outcome

Patient-centered care: Patient-centered care is improved by deep learning, which makes it possible for early detection, accurate diagnosis, and customized treatment planning. Treatment plans tailored to each patient's needs and enhanced patient-provider communication are advantageous. Research Facilitation: By helping researchers analyze large MR image datasets, deep learning algorithms improve their knowledge of the traits, behaviors, and treatment responses of brain tumors. This may promote the creation of innovative treatments and improvements in the field of neuro-oncology research. Empowerment of Healthcare personnel: Deep learning provides advanced diagnostic capabilities and decision support, making it a useful tool for healthcare personnel. This enhances patient outcomes and fortifies their ability to deliver superior healthcare. Decreased Healthcare expenditures: By avoiding needless operations, protracted hospital stays, and therapies associated with late-stage or incorrectly diagnosed brain tumors, early identification and accurate classification of these tumors can lower healthcare expenditures. This could lessen the financial strain on healthcare systems and patients. Research Facilitation: By examining big MR image datasets, deep learning algorithms assist researchers in gaining a better understanding of the traits, behaviors, and treatment responses of brain tumors. This could result in more research being done on neuro-oncology and the creation of innovative treatment plans.

9 Conclusion and Future Scope

This paper provides a summary of various segmentation techniques. This method accurately detects and categorizes brain cancers by segmenting MRI images automatically. by utilizing cutting-edge methods for image processing. Wavelet-based techniques are essential for preprocessing because they remove undesired backgrounds from image, lower noise, and sharpen edges. Our filtering method is distinct from the others since it has the potential to improve both computation efficiency and image quality.

After image enhancement, we applied a magnetic resonance imaging (MRI) segmentation technique that successfully separates the tumor from the surrounding tissue. Although classification-based segmentation produces trustworthy results over a huge dataset, problems may arise if there is insufficient representation of a particular tumor categorization in the training set. These methods determine the type of tumor and establish its presence. After image enhancement, we applied a magnetic resonance imaging (MRI) segmentation technique that successfully separates the tumor from the surrounding tissue. Although classification-based segmentation produces trustworthy results over a huge dataset, problems may arise if there is insufficient representation of a particular tumor categorization in the training set. These methods find tumors and tell what type they are.

In the future, Three-dimensional MRI imaging makes it possible to detect cancers and determine the approximate nature and stage of those tumors, which may lead to more accurate diagnoses and treatment plans. MRIs and CT scans remain the two most widely used methods for the detection of brain malignancies. Medical professionals use physical copies of image to manually identify tumors. Our work proposes a radiograph

scan segmentation method for brain malignancies based on a general-purpose computer system. The recommended system might be utilized on typical machines. The hit ratio for every image is calculated. For scanned images, it attains precise engineering parameter handling. The task that has been done will be compared to earlier projects in order to further boost efficiency.

References

1. Yu, W., Chen, X., Wu, L.: Segmentation of concealed objects in passive millimeter wave images based on the gaussian mixture model. J. Infrared Millimeter Terahertz Waves **36**(4), 400–421 (2015)
2. Kharrat, A., Benamrane, N., Messaoud, M.B., Abid, M.: Detection of brain tumor in medical images. In: Proc. of 2009 3rd International Conference on Signals, Circuits and Systems (SCS), pp. 1–6, November 6-8 (2009)
3. Lundervold, A.S.: An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik **29**(2), 102–127 (2019)
4. Nayak, D.R., Dash, R., Majhi, B., Prasad, V.: Automated pathological brain detection system: a fast discrete curvelet transform and probabilistic neural network based approach. Expert Syst. Appl. **88**, 152–164 (2017)
5. Kavitha, A.R., Chellamuthu, C.: Brain tumour segmentation and detection using modified region growing and genetic algorithm in MRI images. Int. J. Med. Eng. Inform. **9**(3), 269–283 (2017)
6. Kapse, R.S., Salankar, S.S., Babar, M.: Literature survey on detection of brain tumor from MRI images. IOSR J. Electron. Commun. Eng. (IOSR-JECE) e-ISSN: 2278-2834, p-ISSN: 2278-8735. Volume 10, Issue 1, Ver. II (Jan-Feb.2015), pp. 80–86
7. Rasel Ahmmed Anirban Sen Swkshar,Md.Foisal ,and Md.Abdur Rafiq: Classification of Tumors and It Stages in Brain MRI Using Support Vector Machine and Artificial Neural Network. In: International Conference on Eletrical,Computer and Communication Engineering(ECCE),Cox's Bazar, Bangladesh, February 16–18 (2017)
8. Sijbers, J., et al.: Estimation of the noise in magnitude MR images. Magn Reson Imaging **16**(1), 87–90 (1998)
9. Almadhoun, H.R., Abu-Naser, S.S.: Detection of brain tumor using deep learning. Int. J. Acad. Eng. Res. (IJAER) **6**, 29–47 (2022)
10. Musallam, A.S., Sherif, A.S., Hussein, M.K.: A new convolutional neural network architecture for automatic detection of brain tumors in magnetic resonance imaging images. IEEE Access **10**, 2775–2782 (2022) [CrossRef]
11. Wo´zniak, M., Si?ka, J., Wieczorek, M.: Deep neural network correlation learning mechanism for CT brain tumor detection. Neural Comput. Appl. 2021, 1–16. Available online: https://link.springer.com/article/https://doi.org/10.1007/s00521-021-05841-x. Accessed on 15 Nov 2022
12. Amin, J., Sharif, M., Haldorai, A., Yasmin, M., Nayak, R.S. Brain tumor detection and classification using machine learning: a comprehensive survey. Complex Intell. Syst. **8**, 3161–3183 (2022). [CrossRef]
13. Nayak, D.R., Padhy, N., Mallick, P.K., Zymbler, M., Kumar, S.: Brain Tumor Classifification Using Dense Effificient-Net. Axioms **11**, 34 (2022) [CrossRef]
14. Obeidavi, M.R., Maghooli, K.: Tumor detection in brain MRI using residual convolutional neural networks. In: Proceedings of the 2022 IEEE International Conference on Machine Vision and Image Processing (MVIP), Ahvaz, Iran, 23–24 February 2022, pp. 1–5 (2022)

15. Khalil, H.A., Darwish, S., Ibrahim, Y.M., Hassan, O.F.: 3D-MRI brain tumor detection model using modifified version of level set segmentation based on dragonflfly algorithm. Symmetry 2020, 12, 1256. [CrossRef]
16. Sajid, S., Hussain, S., Sarwar, A.: Brain tumor detection and segmentation in MR images using deep learning. Arab. J. Sci. Eng. **44**, 9249–9261 (2019). [CrossRef]
17. Lotlikar, V.S., Satpute, N., Gupta, A.: Brain tumor detection using machine learning and deep learning: a review. Curr. Med. Imaging **18**, 604–622 (2022). [CrossRef]
18. Mahmud, Md.I., et al.: A dseep analysis of brain tumor detection from MR images using deep learning networks. Algorithms **16**(4), 176 (2023). https://doi.org/10.3390/a16040176
19. Madheswaran, M., Anto Sahaya Dhas, D.: Classification of brain MRI images using support vector machine with various Kernels. ISSN 0970-938X, April 2015
20. Hussein, E.M., Mahmoud, D.M.A.: Brain tumor detection using artificial neural networks. J. Sci. Technol. **13**(2), December 2012
21. Abdel-Maksoud, E., Elmogy, M., AlAwadi, R.: Brain tumor segmentation based on a hybrid clustering technique. Egyptian Inform. J. **2015**(16), 71–81 (2015)
22. Sathees, B., Anbu Selvi, R.: Feature extraction using image mining techniques to identify brain tumors. In: IEEE International Conference on ICIIECS'15
23. Kaur, H., Kaur, S.: Improved brain tumor detection using object based segmentation. (IJETT), Volume 13 Number 1, Jul 2014
24. Kaur, A.: An automatic brain tumor extraction system using different segmentation methods. In: 2016 Second International Conference on Computational Intelligence & Communication Technology
25. Gahukar, S.D., et al.: Int. Journal of Engineering Research and Applications Vol. 4, Issue 4(Version 5), April 2014, pp.107–111

A Study on Employee Engagement Factors of Information and Communication Technology (ICT) Workforce

Gowri Ramakrishnan(✉) and Neeraj Kumar Dubey

Amity University Madhya Pradesh, Gwalior, India
gowri.rams@gmail.com

Abstract. Organizations desire to achieve goals to sustain in competitive environment. Crisis like Covid-19, Economic crisis, war, etc. affect Employee Engagement due to sudden changes in organizational environment. Employees' attitude towards work changes and hence Strategic Human Resource Management (SHRM) measures are needed during crisis. Predicting employee engagement is ever challenging for Information & Communication Technology (ICT) employees. Assessment of ICT employees was carried out through online questionnaire survey and analysis using correlation & regression. Results and implications imperative for ICT employees have been cross validated with contextual findings and specific envisaged outcomes. Modern tools are now supported by Natural language processing and character mapping with infusion of Artificial Intelligence (AI) -& Machine Learning ML) can provide SHRM decisions. The results show that the workload, team integration and mode of work has significant relationship with employee engagement. Team integration plays vital role along with the amount of work & mode of working. Social media communications & time with family has no significant relationship. Team-relationship increases engagement & retention.

Keywords: Employee Engagement · Crisis · Talent Retention · Information & communication Technology

1 Introduction

Talent Retention is an ever-challenging task for all types of organizations especially during great resignation waves. Rapid quitting, retirement or laying off by employees in the year 2021 has been rechristened again as one of the Great resignations. Rate of resignation has been non-uniform across the industries. However, attrition studies from April 2021 to Sep 2021 [46] reveal that Industry Average Attrition Rate of 18% to 7% are Information Communication and Technology (ICT) related employees'. Also, seasonally adjusted household data of U.S. Bureau of Labour Statistics [7] for Dec 2021 shows that unemployment rate has reduced to 3.9 (closer to pre-Covid-19 figure of 3.5) from 6.7 a year back. While employment in professional and business services especially in computer system design and related services demonstrate rising trend. These dynamics in competitive & disruptive environment with positive job creation

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 320–337, 2025.
https://doi.org/10.1007/978-3-031-86305-9_22

lacks competent, proficient & talented workforce. Skillsoft's Global Knowledge 2021 IT Skills and Salary Report [53] undertaken globally suggest that salary or compensation which used to be the driving factor traditionally is no more a main driver for employee turnover. In the past three years, opportunity for growth & development has emerged as a key factor for more than half of job switching employees. Also, nearly 31% of these professionals attributed to insufficient work/life balance and demanded reasonable distribution of job requirement & personal quality time. While only 19% connected themselves with company's performance.

Job roles either phased out or changed due to automation or advancement in technologies contribute about 6% indicating that you need to be current and contemporary in the domain to maintain your credible employment. These indicate that Talent retention is still a problem as technically qualified resources like ICT employees. Therefore, Companies continuously train employees to increase their proficiency in contemporary technologies to retain the employees engaged at work and reduce stress [21].

Work culture, job insecurity & work culture, lack of recognition and poor response during crisis are various predicted reasons for great resignation wave [46]. Crisis conditions like Covid-19 restrictions, inflation, war situations, economic uncertainties, etc. affects employees' health physically and emotionally. Also, crisis situations leads to psychological distress which resulted in great resignation wave [23].

Organizational behavior and culture during crisis lead to anxiety, depression and stress. Distress among employees occurs due to various reasons such as bullying at workplace, unfair treatment, poor communication, effort or reward imbalance, role conflicts, lack of appreciation, support, transparency, and authority [5]. Employees productivity and employee commitment are affected by job stress [15]. Lack of time for regular exercise lead to increase in health concerns significantly [44]. In Indian cities, 80% of 6,000 employees have symptoms of anxiety and among them 55% of employees are in depression at 2016 as per the survey conducted by peoplematters India [2]. Working conditions in organization causes psychological distress among employees [29].

Distress directly affects employee engagement and increases employee turnover & reduces quality of work output. Participation or passion of employees towards their job increases their involvement & effective quality of work is a classical way to define Employee Engagement. Cost of training & hiring new resources and timely deployment actively engage in their work impacts organizational productivity & organizational performance. Many researches confirms that employee engagement positively influences talent retention in organization [25, 30]. Researchers also suggest psychometric test or surveys as a primary tool to measure employee engagement. During crisis, new challenges are faced by organization due to dynamics faced by employees and organizations like changes in organizational and social structure due to work from home. Limited research is available on predicting various stressors that affect employee engagement during & after crisis. Prediction of stressors helps to improve employee engagement to retain talent. Block diagram of a typical analysis of impact of stressors in employee engagement and talent retention is shown in Fig. 1.

Therefore, problem statement can be framed as 'Prediction of Employee Engagement of ICT employees (which affect employee and organizational productivity) has found limited application by organizations in designing competitive strategies in this disruptive

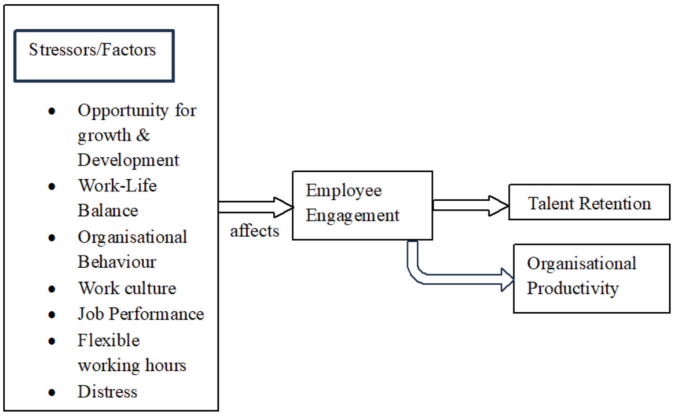


Fig. 1. Impact of Stressors in Employee Engagement and Talent Retention

environment'. With effective changes in external environment and internal environment, optimal utilisation of Employee Engagement is becoming a quintessential key to talent retention.

It is imperative to analyse how organizations or individuals behaved during crisis time. Towards this, dependent and independent factors/stressors that affect the organization & individual have been identified and study of stressors that affect employee engagement has been assessed. Findings will help organizations to focus on those stressors to reduce distress among employees, increase employee engagement and effectively reducing employee turnover. The relationship between various factors or stressors that affect employee engagement will determine strategies in this competitive & disruptive environment.

2 Review of Literature

Majority of workforce who quit their jobs either to take new jobs or to leave is termed as Great Resignation [17]. The term 'Great Resignation' was first coined by Prof. Anthony Klotz in May 2021. Job seekers changed their jobs looking for change in their present profession or for better compensation, challenging positions, work life balance. Research on current workforce indicates that employees would retain in an organization which offers challenging & satisfied jobs, learning & development and engagement [24]. According to Global Knowledge IT skills and salary report of 2021 only 28% of employees wants to take new certification training to prepare for different jobs; while 74% of employees train to improve their skill-set [53]. ICT corporates have employed various strategies to retain their employees. For example, academic interface programme is launched by TCS and offers courses for learning and development, Initial and Continuous Learning Programs, Leadership Development Programs, Foreign Language Initiative, Workplace Learning and Ignite [39]. Research on retention strategies for ICT companies such as TCS, Infosys, Accenture, Wipro and HCL, the most used include Rewards & Recognition to deserved employees, providing Onsite opportunities, employee welfare measures,

Relocation with promotion and career learning & development [12]. Organization could not retain its talent when there is less engagement of employees who needs better organizational culture or values, relationship with team members and leaders, self-development activities [24].

Stress was defined as "The physiological or psychological response to internal or external stressors" by Hans Seyle in the year 1940. Stress can be based on actions or responses of individual for various activities, events both by means of internal or external environment. Stressors are stress triggering events. Stress levels may vary depends on the handling or acceptance of individual. Responses which are positive for stressful events are called as Eustress. Eustress helps to improve the performance of individual. The increase in continuous stress will results in Distress which is negative impact of stressful event [37]. Research on job stress implies that employees become distress because of organization [29] and it is explained by Beehr and Newman in the year 1978 as "job stress as a condition where in job related factors interact with the worker to change (disrupt or enhance) his or her psychological or physiological condition such that the person (mind or body) is forced to deviate from normal functioning" (Beehr & Newman, 1978).

2.1 Employee Engagement Concepts & Models

Employee engagement (EE) was introduced as 'personal engagement' was that 'harnessing of organization members' selves to their work role [19]; later EE influenced the traditional concept of HRM to positive HRM which often dictated performance of individual & organization [48]. Since then, Employee engagement has been drivers for formulation of best practices of HRM to mitigate the implementation challenges of Strategic HRM. These foundation researches on EE guided organizations to increase their performance by practicing employee engagement initiatives. Most researches are based on numerical assessment of engagement data obtained only after employing initiatives. This data is applied to determine workplace and environment experiences [35]. However, during crisis situations both internal and external environment affects employee engagement. Crisis can be defined as situation in which either or both the individual and the organization is subjected to restraints due to internal or external operating environment. Necessary change in EE strategies and timely application can achieve planned organizational performance. Studies on Covid-19 dynamics in the workplace environment reveals that even factors external to work landscape can influence employee engagement varying with individual identities [14]. The various techniques and elements to measure EE under various business environments are:-

1. Hewitt Engagement Model measured three paradigms such as "Say, Stay and Strive" by employee survey. This survey include leadership, pay, work balance, and empowerment [40].
2. Factors like Task performance, Organization citizenship behavior, Counterproductive/deviant workplace behavior, Trait affect, Employee engagement, Job satisfaction, Organizational commitment, Perceived organizational support, Work centrality and Job involvement were ascertained using different standard questionnaire (both pen-pencil mode and online survey method) and univariate & multivariate weight

analysis method to assess data. Trait affect determines job satisfaction and employee engagement. Also proper measurement of Employee engagement ensures prediction of cognitive-affective of Organization Citizenship Behavior [9].

3. Factors such as work environment, team, co-worker relationship, training and career development, compensation programme, policies & procedures and workplace well-being, employee productivity (EP) and employee engagement were measured using questionnaire survey method with five-point Likert scale (ranging strongly agree to strongly disagree). It found that work environment, team, co-worker relationship played important role in employee engagement [18].
4. External environment impacts employees' way of work at office. The focus group method was used to collect the survey. Factors such as geographical location, relationship outside the work, promotion, learning new technologies determine employees engagement at work [26].
5. Employee engagement of work from home employees is determined using the review of various articles during recent pandemic as a secondary data only [8].
6. Questionnaire survey method through email was carried out to determine EE of WFH employees through various determinant factors such as Leadership, Company practices, Job Security - Safety, Empowerment, Rewards - Recognition, Trust, Technological Readiness and Learning Support. SPSS software is used to calculate impact of various factors. This revealed that two important factors that positively contributes EE for WFH employees are Technological Readiness and Empowerment while the length of service remained as a moderating factor [28].
7. Mixed match approach of structured review of literature, content analysis of work place design & management monitoring tools, employee engagement metrics measuring attitude of employees, feelings of employees, work behavior and social relationships were determined as determinants [47].

2.2 Employee Engagement Approaches in ICT Companies

Most organizations have been forced to switch over to remote working due to government protocols contemplating the welfare of employees during recent pandemic. These changes were effected within inadequate timeframe without analysing the fundamental management principles. With work from home & other modes of work becoming the preferred modes of employment, stressors are likely to impact employee engagement, employee productivity and organizational productivity especially in ICT related sectors. At the same time, work from home which is more prevalent in ICT companies poses growing challenges such as physical and psychological fatigue due to continuous contact with work, lack of attention due to virtual meetings, distraction due to kids, loss of creativity or productivity due to reduction of interaction amongst team members and reduction of mentoring [11]. The following are some of the recent researches with critical comparison of relevant models/methodologies pertaining to employee engagement for talent retention in ICT related companies:-

1. Compared different IT companies' talent retention strategies by questionnaire survey based on 15 retention strategies conducted among employees of 5 top IT companies among Noida region in Delhi and comparison of mean scores & analysis of variance techniques to find the most used retention technique [12].

2. Talent management in Russia was carried out by taking IT Industry as case study [22]. Questionnaire survey method is used to calculate employee attitude.
3. Impact of Talent Management on Employee Engagement, Retention and Value Addition in achieving Organizational Performance [30] studies about the relationship between employee engagement and talent management activities. Retention relationship with the organizational performance is derived by using review of various literature.
4. The systematic literature review was utilized to find out the various methodologies, frameworks of employee engagement to increase emotional commitment and enthusiasm towards HR management practices [51].
5. The structured 5 item scale questionnaire was used to calculate the leadership style, rewards & recognition, training & development and innovation [10].

Retention or Turnover has always been a top priority for HR managers. This is evident by the survey carried out by Society for Human Resource Management (SHRM) with Globoforce [55] which says that Retention or Turnover has peaked as key challenge as envisaged by HR top workforce. From mere 25% in 2012, it climbed to 47% in 2017 making it the shows topper in business environment. Although challenges for employee engagement decreased from 41% to 31%, there have been significant retention issues. This indicates that Employee Engagement is a principle factor that affects turnover. Hence, employee engagement has to be given high priority to retain employees of the organization.

Therefore, it is necessary to determine employee engagement to predict distress to the organization & employee. Employee engagement has been measured using standard work engagement scales such as 9-item Work Engagement Scale [41] which utilises depletion (negative emotional response from employees) and enrichment (positive state of employees), 13-item Engagement Scale [13], 17-item Utrecht Work Engagement Scale [43] which focuses on Activation, Identification and Absorption to measure burnout and engagement, and 11-item Job Engagement Scale [42] which focuses on job satisfaction and job characteristics.

Usage of internal and external social media by employees in organizations has become principle mode for communication especially during crisis. This has caused stress by virtue of usage of technology (technostress) for either proper or improper communication. The increase in technostress impacts EE and organizational commitment [38]. Although, social media has helped employees to adapt to WFH mode, there has been increase in technostress and strain due to excessive usage [34]. Also, time spent to organise and assist meetings increased; while the collaboration with team members have become limited in WFH [16]. This has affected the effective EE and in turn productivity of employees and organization.

3 Methodology

3.1 Hypothesis of the Study

In order to find the impact of factors which affects employee engagement the following hypotheses are formulated for testing.

Relationship Between Workload & Employee Engagement. Workload is not new but still exists. Unexpected events, environmental stressors can affect workload and employee experience [49]. More than 60% of over workload employees are in at risk of turnover and it has effect on their physiological and psychological wellbeing [20]. When the workload increases the performance of the employee decreases. But this may not be completely true in crisis. Therefore, the hypothesis developed with reference to workload is as follows:

H_1: Workload affects Employee Engagement

Relationship Between Team integration & Employee Engagement. Support from supervisor and rewarding co-worker provides psychological meaningfulness to employee which helps in employee engagement [13]. Individual engagement and their performance depends on team engagement [1]. Leaders who supports personal development of team members and celebrates team successes achieve employee engagement [50]. Team dynamics and leadership impacts workforce engagement levels differently in disruptive environment. Hence, the following hypothesis with reference to Team integration is developed as:

H_2: Team integration affects employee engagement.

Relationship Between Distraction due to Family & Employee Engagement. Employees engagement is affected when work life balance does not exist. While this has been a major impediment with mental wellbeing taking positive impact on individual's performance. This has taken on a special significance with the advent of remote working. Stress induced due to work adversely affects home life, personal wellbeing & relationships. This has impacted 88% Indian workforce even to trade their high paying jobs for want of better mental health [54]. This factor has been christened as a negative effect with the following hypothesis:

H_3: Distraction due to family affects employee engagement.

Relationship Between Social Media Implications & Employee Engagement. Now a days social media plays important role with employees and organization. It helps to share information, solve problems, avoid duplication, increase agility in business and advantages of employee engagement [33]. Social media used within organizations helps both current and future potential employees to engage [36]. Yet, the speculation of whether social media has profound impact on employee needs introspection. This has induced the deduction of the following hypothesis:

H_4: Social media implications affect employee engagement.

Relationship Between Mode of work & Employee Engagement. Working from home and mutually integrating with office work during remote conditions can be challenging and loaded with tension [32]. Performance of employees and team is better when they work with their colleagues in office together and individual employee performance is negatively related with work from home [27]. As per the report of Ipsos says 37% of employees feel disengaged from work and 33% feel burned out when they work from home [6]. Remote working has mixed results with policy restrictions imposed by Government and later being adopted as a norm by organization due to selective mutual benefits with employees. Hence, the following hypothesis is imperative:

H_5: Mode of work affects employee engagement.

3.2 Variables of the Study

Independent Variables are various predictors such as workload (WL), Team Integration (TI), Distraction due to Kids (DK), Social Media implications (IM) & Mode of Work (MW). These predictors were selected after extensive literature survey that affects employee engagement. Dependent Variable will be Employee engagement.

3.3 Sample

The objective of sampling design is to assess the factors of Employee Engagement of employees of IT sector. Direct employment in the IT services and BPO/ITeS segment is estimated to reach 4.47 million in FY 2020-2021 (E) with an addition of 1,38,000 people (consisting of ~ 36 percent women employees)[52].

Sample size has been calculated for a confidence level of 95%, acceptable error and precision of 5% for a random sample selection from the IT sector population of about 4.47 million, which is approximated to 73 (using "Small Sample Techniques," the research division of the National Education Association). The sample size has been calculated by the formula published in "Small Sample Techniques," the research division of the National Education Association is mentioned below

$$n = \frac{z^2_{\alpha/2}.p.q.N}{e^2(N-1) + z^2_{\alpha/2}.p.q}$$

where, p = Sample proportion, $q = 1 - p$; $z_{\alpha/2}$ = The value of the standard variate at a given confidence level and to be worked out from table showing area under Normal curve; n = Size of sample; N = Population size; e = acceptable error; Here we have taken the value of $p = 0.95$ so that n will yield maximum sample at the desired precision which will be the most conservative size of sample.

The parameters of interest was includes as the list of identified predictors based on literature review on the determination of Employee Engagement which was further translated in the form of an online questionnaire. Questionnaire was floated on online mode towards collection of sample for duration of 27 days. For this study, the samples selected were 75 employees working in ICT related organizations with diversified cultural background indicating the demographic distribution as per Fig. 2.

3.4 Research Design

Research method for conducting the research is both quantitative inferential approach and qualitative research for validation which investigates the reasons for variance in Employees during crisis. Quantitative inferences based on inputs from employees are studied and the reasons for predictors are analysed deeply. Qualitative methods used to validate the dependencies of predictors to Employees engagement levels. Flowchart of the research method used is as per Fig. 3. The following are key steps and vital design precepts of the research methods:-

1. Sampling design used is random sampling

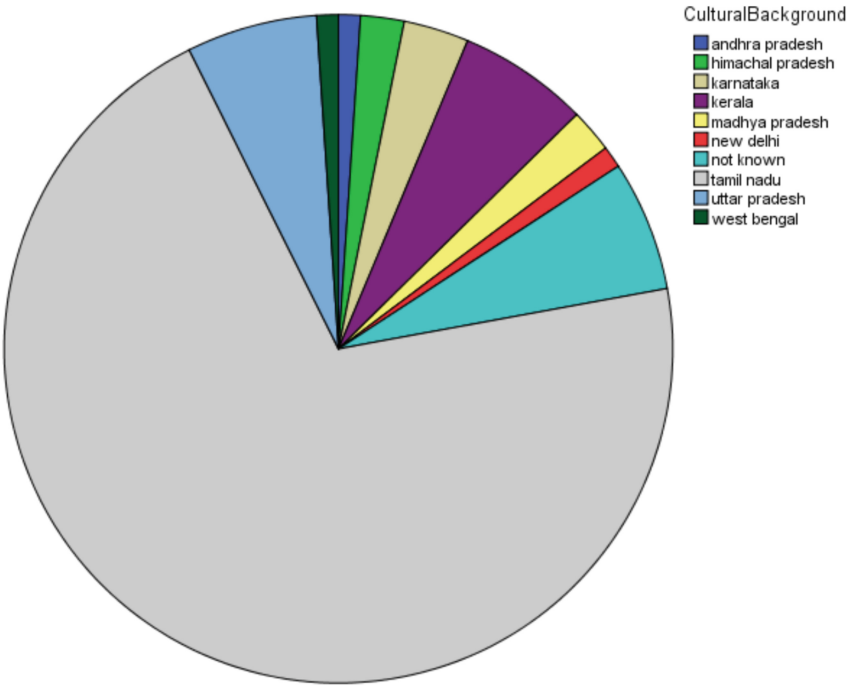


Fig. 2. Cultural Background of sample population

2. Statistical design used is Subjective & analytic analysis for reduction of the factors or predictors & standard statistical tools utilised to ascertain results.
3. Online structured questionnaire survey was used for collecting data from working employees of private and public organizations who works in work from home, work from office and hybrid mode.

3.5 Test/Tools Used for Data Collection

SPSS tool was utilised to carry out analysis of the data collected through the following modes:-

1. Online questionnaire is based on 5-Predictor analysis with 30-Item questions against a 6-point satisfaction scale with the descriptors – Dissatisfied (DS), Partly dissatisfied (PDS), Satisfied (S), Partly satisfied (PS), Fully satisfied (FS) & Extremely satisfied (ES). Table 1 shows number of statements in questionnaire. Questions on Gender, Size of organization, Mode of work, Type of employment, Total number of experiences in current organizations are designed. Interview method was not required as all respondents are provided with adequate information.

2. Secondary data from various relevant valid sources were utilized to validate results.

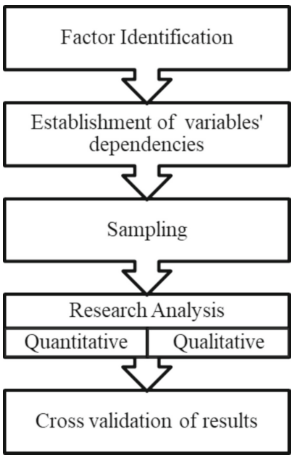


Fig. 3. Flowchart of Research Method

Table 1. Number of statements under each predictor in questionnaire

S. No.	Factors / Predictors	Statements	Total
1	Workload	1, 3, 4, 6	4
2	Team Integration	8, 9, 10, 12	4
3	Distraction due to kids	13, 15	2
4	Social Media Implications	17, 18, 19, 20	4
5	Mode of work	21, 22, 23, 24, 25, 26, 27, 29	8
6	Employee Engagement	2, 5, 7, 11, 14, 16, 28, 30	8

4 Results and Discussion

Primary data collected through online questionnaire are scrutinized for valid responses in terms of completeness and authentication. Data analysis for this quantitative study was started with assigning numerical value to descriptors as shown in Table 2.

Table 2. Assigning values to descriptors

Items	Dissatisfied (DS)	Partially Dissatisfied (PDS)	Satisfied (S)	Partially Satisfied (PS)	Fully Satisfied (FS)	Extremely Satisfied (ES)
Statements	1	2	3	4	5	6

The means of assigned values to various questions on work load, team integration, distraction due to kids, social media implications, mode of work were calculated

using Statistical Package for Social Science (SPSS) program. Pearson's product moment correlation were calculated which shows the following results as shown in Table 3.

Table 3. Pearson Correlation Results

S. No.	Factors/stressors	Employee engagement
1	Workload	0.730
2	Team integration	0.840
3	Distraction due to kids	0.708
4	Social media implications	0.645
5	Mode of work	0.731

From this we can interpret that predictor team integration has highest value of 0.840 in relation with employee engagement. This shows that recognition from the team leader, working as a team, mutual cooperation between team members in terms of learning and development, helping each other in their work to attain better results helps in increasing employee engagement.

Workload factor attains the next highest relation with employee engagement. Completion of work within the given time, maintaining work life balance, taking responsibility of work provides more employee engagement than working for long hours.

Remote working considered as a predictor for this study revealed to be significant in terms of the effect on employee engagement as well. Optimization of work by meeting timelines, amount of work given, office organized fun activities, virtual meetings attended are having high correlation with employee engagement and it shows less significance in going back to office after crisis conditions.

Spending time with family and relaxing with kids increases employee engagement. Finally social media implications, the social media communication tools such as google meet, cisco Webex etc. and internal social media communications in organizations helps to improve employee engagement while the interruption by social media and creativity of employee has less significance in employee engagement.

Table 4. Regression Analysis Model Summary

Model	R	R Square	Adjusted R square	Std. Error of the Estimate	Change Statistics				
					R Square Change	F Change	df1	Df2	Sig. F Change
1	.908	.825	.812	.42115	.825	65.073	5	69	.000

Regression model (Table 4) summarizes the following:

1. R-value is 0.908 which is greater than 0.4 represents the strong correlation between dependent and independent variable. Hence further analysis can be carried out.

2. R-Square value represents the total variation of dependent variable due to independent variables. R-Square value shows that this model is sufficient to determine the relationship as 0.825 which is greater than 0.5.
3. Adjusted R-Square provides generalisation of outcomes. This indicates variation of sample results from population in multiple regression. Here the difference between R-Square (0.825) and Adjusted R-Square (0.812) is only 0.013 which indicates that both the values are close and not far off.

Bivariate regression analysis carried out on the processed data shows that strength of relationship in terms of the significance of variable and magnitude of impact on dependent variable. The following Table 5 shows the analysis carried out to find the hypothesis testing for the study undertaken:-

Table 5. Regression Co-efficient Values and its Significance with Employee Engagement

Independent Variable	Sig. Value	Hypothesis Testing Result at 95% of confidence level	Interpretation
Workload	0.004	Null Hypothesis Rejected ($0.004 < 0.05$)	There is a significant relationship between workload and employee engagement
Team Integration	0.000	Null Hypothesis Rejected ($0.000 < 0.05$)	There is a significant relationship between Team Integration and employee engagement
Distraction Due to Kids	0.226	Null Hypothesis Accepted ($0.226 > 0.05$)	There is no significant relationship between distraction due to kids and employee engagement
Social Media Implications	0.836	Null Hypothesis Accepted ($0.836 > 0.05$)	There is no significant relationship between social media implications and employee engagement
Mode of work	0.021	Null Hypothesis Rejected ($0.021 < 0.05$)	There is significant relationship between mode of work and employee engagement

The co-efficient table signifies the acceptance level based on significant value to either reject or not rejected. The above table indicates that all independent factors of the study – Workload, Team Integration, Mode of work has significant relationship with employee engagement except Social Media Implications and Distraction Due to Kids.

The employee engaged in the work provides their best results to organization. The above findings have been validated by secondary data as follows:

1. Work Load: The intermediate work i.e., work finished within a given time increases engagement among employees and also increases the innovation among workers [31]. High workload leads burnout among employees and in turn affects employees engagement. Our survey results show that workload affects employee engagement (H_1) is true. There is a significant relationship between workload and employee engagement which is congruent to the secondary data.
2. Team Integration: Team integration plays a vital role in employee engagement. The leadership in team should recognize talent and provide appreciation to all team members. The team which is supported by the leader has more team spirit and shows effective employee engagement. [50]. Our study results also show that (H_2) team integration affects employee engagement. There is a significant relationship between team integration and employee engagement which is still valid in 2022 post crisis condition.
3. Distraction due to kids: Spending time with family and kids which is otherwise called as work life balance does not have significant relationship with employee engagement. The organizations which provide employees flexible time based working to take care of the dependents of the employee and providing personal leave helps in the success of engagement of employees [3]. However, our survey results show that distraction due to kids does not affect employee engagement (H_3). There is no significant relationship which is less discernible because presently the preferred mode of work is either work from home or hybrid work implying the perceived influence of distraction due to kids.
4. Social media implications: The employees spending time in internal social media of organizations, external social media applications and communication tools. Our survey shows that social media implications does not affect employee engagement (H_4). A similar study in a Telecom company [36] also revealed that social media usage will not be helpful in engaging employees unless there is open communication and participation in culture and leadership of the organization.
5. Mode of work: Organizations started providing hybrid work, work from office and work from home as per the requirement of the team. Our survey results shows that mode of work affects employee engagement (H_5). Employees productivity increased during remote working and most of the employees are choosing for hybrid work [45].

5 Conclusion and Implications

5.1 Conclusions

Employee engagement helps to improve employee performance, employee productivity and organizational productivity. Organizations are finding ways to improve employee engagement in terms of providing recognition, learning and development of new concepts or new technologies, organizing team building activities, flexibility in working time. While the crisis conditions in this disruptive environment such as Covid-19, Conflicts among nations, economic slowdown, famines, etc. make the employees to prioritize their family and eventually organizations have started implementing new strategies for better engagement of employees. These crisis situations also pose new challenges to modern organizations and forcing them to adapt to changing policies and regulations.

The primary objective of this research revealed that the dependent factor as employee engagement which was impacted by independent factors such as type of employment, workload, team integration, distraction due to kids, social media implications. They affect organization and individuals significantly.

Pearson correlation coefficient and Bivariate regression techniques were used to analyze the results on self-prepared online questionnaire-based survey. Results indicate that predictors such as mode of work, workload, team integration has significant relationship with employee engagement except distraction due to kids, social media implications. These results are found to be congruent by validation against secondary data from articles and research papers which is indicative of the fact that this study carried out is successful.

The study also deciphered that the principle factor such as Team dynamics which directly impacted engagement levels of general populace of any organization. Sense of ownership and mindfulness & commitment towards task often depended on Team rather than individual. But during crisis, as the team dynamics could not be established, the individual actions led to disengagement from organization goals. Post crisis, with the resumption of contact with team ultimately connected with the objectives. This reveals that human behavior is directly impacted by engagement levels and have significant effects of psychological equation with Organizational behavior. Team building activities builds social relationships among employees which helps in mutual learning and understanding, reduces work and mental stress and reduces attrition rate in organizations.

Inadequacy of time to complete work directly affects employees' which makes them feel dissatisfied at job. Optimization of work-time, work-load & staffing can accentuate even innovation culture by maintaining engagement levels of individuals. Flexibility in work offers alternative approach to hard bound timelines and provides time & space to be managed by individual to produce results in a near optimal way.

5.2 Implications

This research provides various implications to managers and leaders of organizations which include recognition by Team Leaders; mutual helping tendency by all team members in terms of learning and development of new technologies; work load optimization to provide quality time with family; and finally the flexible way of working helps to improve psychological conditions of employees to remain engaged.

In practice, policies and regulations must be flexible especially during crisis and it should focus on boosting the open communications, social interactions among team members to increase engagement of employees, employee productivity and organizational productivity.

In Future, all employees should be given adequate time for finishing their jobs to maintain optimum stress levels for the growth of individual & organization. Employees should be recognized for their work and provided adequate appreciation in a planned frequency. Finally, flexibility in working conditions needs be provided to the employees especially during crisis & post crisis conditions to enable recovery to normalcy.

This research aids in design of advisory for HR managers, guidelines for employees and priority list for organizations. Typical use-cases include formulation of employee

engagement initiatives based on entry & exit interview questionnaire and assessment of efficacy of existing employee engagement & talent retention measures.

6 Delimitations, Limitations & Future Suggestions

This study is delimited to employees who are having internet access only as online questionnaire was used for data collection. This pilot study is limited by the sample size and demography, therefore a detailed sampling needs to be undertaken for application.

During the analysis, the mean of response was used to recode for grouping various indicators of Factor/Predictors/Stressors and then various techniques are applied on this recoded data which may have resulted in error migration. However, the validation using the secondary data indicated that the study undertakes was not in significant deviation with contemporary study results. Although, option of more factors could have been applied, only limited factors have been analysed to orient the study to meet specific objectives.

Target industry could be any of the industries such as education, banking etc. However, this study should be restricted to a specific industrial sector for detailed research and analysis. The future studies could compare employee engagement models across different sectors or within same sector between corporate and startup companies. This provides accurate results for specified industry. Additionally, this would help the organizations to design an engagement-based framework for implementation. This should be periodically reviewed using survey questionnaire/interviews. The face-to-face interview can be conducted to observe the validate the responses of the respondents.

Declarations.

Conflict of Interest All authors certify that there is no actual or potential conflict of interest in relation to this manuscript.
Informed Consent All authors certify that privacy of individual is not disclosed in this manuscript.

Funding Information. All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Ethical Approval. All authors certify that manuscript is not submitted to multiple journals for simultaneous consideration.

References

1. Bakker, A.B., et al.: Key questions regarding work engagement. Eur. J. Work Organ. Psychol. **20**(1), 4–28 (2011). https://doi.org/10.1080/1359432X.2010.485352
2. Banaji, V.: Article: Corporate India's mental health crisis — People Matters, https://www.peoplematters.in/article/wellness/corporate-indias-mental-health-crisis-24337. Accessed 29 Nov 2021
3. Bedarkar, M., Pandita, D.: A study on the drivers of employee engagement impacting employee performance. Procedia - Soc. Behav. Sci. **133**, 106–115 (2014). https://doi.org/10.1016/j.sbspro.2014.04.174

4. Beehr, T.A., Newman, J.E.: Job stress, employee health, and organizational effectiveness: a facet analysis, model, and literature review. Pers. Psychol. **31**(4), 665–699 (1978). https://doi.org/10.1111/j.1744-6570.1978.tb02118.x
5. Bhui, K., et al.: Perceptions of work stress causes and effective interventions in employees working in public, private and non-governmental organisations: a qualitative study. BJPsych Bull. **40**(6), 318–325 (2016). https://doi.org/10.1192/pb.bp.115.050823
6. Boyon, N.: Back to the workplace (2020)
7. Bureau of Labor and Statistics: Employment Situation Summary - 2021 M13 Results. https://www.bls.gov/news.release/empsit.nr0.htm. Accessed 28 Jan 2022
8. Chanana, N.: Sangeeta: Employee engagement practices during COVID-19 lockdown (2020). https://doi.org/10.1002/pa.2508.
9. Dalal, R.S., et al.: The relative importance of employee engagement, other job attitudes, and trait affect as predictors of job performance. J. Appl. Soc. Psychol. **42**(SUPPL. 1), 295–325 (2012). https://doi.org/10.1111/j.1559-1816.2012.01017.x
10. Darwin Joseph, R., et al.: Drivers of Employee Engagement andInnovation in Information Technology Industry. **20**, 38–46 (2018). https://doi.org/10.9790/487X-2001083846
11. Deloitte: Remote Collaboration Facing the challenges of COVID-19. Deloitte. March, 1–13 (2020)
12. Dhillon, M.: Employee Retention Strategies - in Indian Information Technology Sector. Int. J. Manag. 11, 12, 1621–1629 (2020). https://doi.org/10.34218/ijm.11.12.2020.148.
13. Douglas, R., et al.: The psychological conditions of meaningfulness, safety and …. J. Occup. Organ. Psychol. **77**, 11–37 (2004)
14. E. Kobena Osam, Nicole Dillard, D.P.: Engagement in times of crises 1 Engagement Under Pressure: The Impact of Identity on Engagement During Times of Crisis. Online J. Work. Educ. Dev. 11, 1, 1–23 (2021)
15. Ekienabor, E.: Impact of Job Stress on Employee's Productivity and Commitment. Int. J. Res. Business, Manag. Account. **2**(5), 124–133 (2019)
16. Gibbs, M., et al.: Work from Home & Productivity: Evidence from Personnel & Analytics Data on IT Professionals. SSRN Electron. J. (2021). https://doi.org/10.2139/ssrn.3843197
17. Hobijn, B.: "Great Resignations" Are Common During Fast Recoveries. Res. from Fed. Reserv. Bank San Fr. 2022, 08, 1–6 (2022)
18. J., A.: Determinants of employee engagement and their impact on employee performance. Int. J. Product. Perform. Manag. 63(3), 308–323 (2014). https://doi.org/10.1108/IJPPM-01-2013-0008.
19. Kahn, W.A.: Psychological conditions of personal engagement and disengagement at work. Acad. Manag. J. **33**(4), 692–724 (1990). https://doi.org/10.5465/256287
20. Kennell, R.: Too Much, Too Little When Workload Affects Wellbeing [white paper]. (2021).
21. Knowledge, G.: 10 Benefits of IT Certification for You (and Your Employer)|Global Knowledge, https://www.globalknowledge.com/us-en/resources/resource-library/articles/10-benefits-of-it-certification-for-you-and-your-employer/. Accessed 28 Jan 2022
22. Kostina, A., Yakushkina, T.: Talent Management : Generalized Approach for IT Industry in Russia. 1–19 (2020)
23. Ksinan Jiskrova, G.: Impact of COVID-19 pandemic on the workforce: from psychological distress to the Great Resignation. J. Epidemiol. Community Health. **76**(6), 2021–2022 (2022). https://doi.org/10.1136/jech-2022-218826
24. Kuzior, A., et al.: Great Resignation—Ethical, Cultural, Relational, and Personal Dimensions of Generation Y and Z Employees' Engagement. Sustainability. **14**(11), 6764 (2022). https://doi.org/10.3390/su14116764.
25. Kyndt, E., et al.: Employee retention: organisational and personal perspectives. Vocat. Learn. **2**(3), 195–215 (2009). https://doi.org/10.1007/s12186-009-9024-7

26. LaMotte, S.: Employee Engagement Depends on What Happens Outside of the Office. Harvard Bus. Rev. Digit. Artic. 2–5 (2015)
27. Lippe, T., Der, Z., L.: Co-workers working from home and individual and team performance. New Technol. Work Employ. **35**(1), 60–79 (2019)
28. Manjaree, H.M.S., Perera, D.A.S.: Determinants of Employee Engagement during COVID 19 Pandemic (Case of Sri Lanka Technological Campus). Int. J. Res. Innov. Soc. Sci. 05, 07, 773–778 (2021). https://doi.org/10.47772/ijriss.2021.5743.
29. Marchand, A., et al.: Does work really cause distress? The contribution of occupational structure and work organization to the experience of psychological distress. Soc. Sci. Med. **61**(1), 1–14 (2005). https://doi.org/10.1016/j.socscimed.2004.11.037
30. Mohammed, A.: The Impact of Talent Management on Employee Engagement, Retention and Value Addition in achieving Organizational Performance. Int. J. Core Eng. Manag. **1**(12), 9510 (2015)
31. Montani, F., et al.: Examining the inverted U-shaped relationship between workload and innovative work behavior: The role of work engagement and mindfulness. (2020). https://doi.org/10.1177/0018726718819055
32. Mulki, J., et al.: Set up remote workers to thrive. MIT Sloan Manag. Rev. **51**(1), 63–69 (2009)
33. Murphy, G.: QUT Digital Repository : This is the published version of the following conference paper : Murphy , Glen D . (2010) Using Web 2 . 0 tools to facilitate knowledge transfer in complex organisational environments : a primer . In : ICOMS Asset Management Co. Eng. Aust. June, 0–14 (2010)
34. Oksanen, A., et al.: COVID-19 crisis and digital stressors at work: a longitudinal study on the Finnish working population. Comput. Human Behav. **122**, 106853 (2021). https://doi.org/10.1016/j.chb.2021.106853
35. Osam, K., et al.: Happiness and healthiness: a replication study. Hum. Resour. Dev. Q. **31**(1), 75–89 (2020). https://doi.org/10.1002/hrdq.21373
36. Parry, E., Solidoro, A.: Social media as a mechanism for engagement? Emerald Group Publishing Limited (2013). https://doi.org/10.1108/S1877-6361(2013)0000012010
37. Quick, A., et al.: Preventive stress management in organizations. (2013). https://doi.org/10.5860/choice.50-3366
38. Ragu-Nathan, T.S., et al.: The consequences of technostress for end users in organizations: conceptual development and validation. Inf. Syst. Res. **19**(4), 417–433 (2008). https://doi.org/10.1287/isre.1070.0165
39. Raheja, S.: An Empirical Study of Selected Indian Companies ' Talent. **1**, 1 (2021)
40. van Rooy, D.L., et al.: Measuring employee engagement during a financial downturn: business imperative or nuisance? J. Bus. Psychol. **26**(2), 147–152 (2011). https://doi.org/10.1007/s10869-011-9225-6
41. Rothbard, N.P.: Enriching or depleting? the dynamics of engagement in work and family roles. Adm. Sci. Q. **46**(4), 655–684 (2001). https://doi.org/10.2307/3094827
42. Saks, A.M.: Antecedents and consequences of employee engagement. J. Manag. Psychol. **21**(7), 600–619 (2006). https://doi.org/10.1108/02683940610690169
43. Schaufeli, W.B., Bakker, A.B.: Utrecht work engagement scale Preliminary Manual Version 1.1. Occup. Heal. Psychol. Unit Utr. Univ. December, 1–60 (2004). https://doi.org/10.1037/t01350-000.
44. Sidhu, A.K., et al.: Job stress and its impact on health of employees: a study among officers and supervisors. J. Manag. Dev. **39**(2), 125–144 (2020). https://doi.org/10.1108/JMD-01-2019-0004
45. States, U., Lists, M.: It ' s time to reimagine where and how work will get done Going back to the office is not an easy call to make (2021)
46. Sull, D.: Toxic Culture is Driving the Great Resignation. https://sloanreview.mit.edu/article/toxic-culture-is-driving-the-great-resignation/. Accessed 28 Jan 2022

47. Surma, M.J., et al.: Assessing employee engagement in a post-covid-19 workplace ecosystem. Sustain. **13**, 20 (2021). https://doi.org/10.3390/su132011443
48. Truss, C., et al.: Employee engagement, organisational performance and individual well-being: Exploring the evidence, developing the theory. Int. J. Hum. Resour. Manag. **24**(14), 2657–2669 (2013). https://doi.org/10.1080/09585192.2013.798921
49. Wickens, C., and B.H.: WORKLOAD Implications for Individual. Natl. Res. Counc. 54–55 (1993)
50. Xu, J., Thomas, H.C.: How can leaders achieve high employee engagement. Leadersh. Organ. Dev. J. **32**(4), 399–416 (2011). https://doi.org/10.1108/01437731111134661
51. Yadav, R., Sushil, D.S.: Impact of human resource management practices on employee engagement in it sector in India. Asian J. Manag. Commer. **5**(1), 123–128 (2024). https://doi.org/10.22271/27084515.2024.v5.i1b.247.
52. Employment Generation | Ministry of Electronics and Information Technology, Government of India. https://www.meity.gov.in/content/employment. Accessed 05 Dec 2021
53. Global Knowledge 2021 IT Skills And Salary Report 2021. Ski. Glob. Knowl. (2021).
54. Mental Health at Work: Managers and Money (2023)
55. SHRM/Globoforce Using Recognition and Other Workplace Efforts to Engage Employees. https://www.shrm.org/hr-today/trends-and-forecasting/research-and-surveys/Pages/employee-recognition-2018.aspx. Accessed 29 Jan 2022

A Simulated Manufacturing Study on Rope Size and Number of Splits on the Productivity While Applying the Simplified Drum-Buffer-Rope Methodology

Jagannathan Sekkizhar(✉) and Nagarajan Vivek

PSG Institute of Management, PSG College of Technology, Coimbatore 641004, India
{sekkizhar,vivek}@psgim.ac.in

Abstract. The Theory of Constraints (TOC) body of knowledge has created solutions for many different business scenarios from manufacturing, to project management, to supply chain, to retail, to even new product design. The oldest tool was the Drum-Buffer-Rope (DBR). A simplified version of this was proposed By Eli Schragenheim which assumed the market to be the constraint and used only one buffer. This method is striking for its simplicity of application. One key point missing in the TOC literature is the size of the buffer and its impact on the throughput of the system. Also missing is the impact of the moderating impact of the splitting of the orders received on the relationship between the buffer size and the throughput. This paper proposes two hypotheses that connect buffer size and order split to the throughput.

Keywords: Traditional DBR · SDBR · CCR · rope · resource loading · throughput · profitability · customer demand · POOGI

1 Introduction

Theory of constraints (TOC) was a new approach proposed by Dr Goldratt (Goldratt & Cox, 2013 and Goldratt & Cox, 2016) [3, 4] as a practical approach to manufacturing projects. The main goal was to improve flow of material within the system, without causing any chaos in the system. TOC techniques increase throughput and reliable delivery while decreasing work in progress inventory [1, 5, 6] (Aggarwal, 1985; Johnson, 1986; Koziol, 1988). TOC based manufacturing systems have shown good results in various industries in different countries [7] (Panizzolo, 2016). Goldratt had coined the term Drum-Buffer-Rope (DBR) to denote this system and the term traditional DBR coined by Schragenheim & Dettmer (2000) [8]. The goal of TOC is to achieve a seamless production flow (McCleskey, 2020) [7]. The major enabler for achieving this production flow was summarised as the focusing steps.

This system was based on the five focusing steps listed below

1. Identify the capacity constrained resource (CCR) in the system. This will decide the drum beat of the entire system

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 338–347, 2025.
https://doi.org/10.1007/978-3-031-86305-9_23

2. Exploit this CCR, make sure that it is not idle for any reason
3. Subordinate all other resources to the working of the CCR
4. Elevate the capacity of the CCR to match new customer requirements
5. Avoid inertia or complacency and go back to step 1, repeat the entire cycle.

This system was also called Process of ongoing improvement (POOGI) and it had the following presuppositions

1. Every system usually has a resource that is the limiting factor to the system's production. This resource is called the capacity constrained resource (CCR)
2. The CCR decides the flow of material in the system
3. All other resources will necessarily have more idle time than this resource
4. Any stoppage on this CCR will impact the output of the entire system
5. An elevation of the capacity of the CCR, could lead to a shifting of the CCR status to other resources in the system
6. Individual resources remaining a CCR is easier to deal with when compared to a condition where the system has interactive constraints, meaning a change in one resource affects the capacity of the other resource.

This system proposed three buffers, the CCR buffer, the shipping buffer and the a ssembly buffer, along with three ropes, namely the constraint rope, the shipping rope and the assembly rope. The shipping rope is the customer demand schedule. In addition to this the CCR itself had a schedule to decide which job to do and for how much time. This triggered the release of materials into the system, a process called the constraint rope. The important objectives of this whole setup was

1. Avoid getting materials too early into the system
2. Avoid overloading the CCR
3. Avoid starving the CCR
4. Ensuring delivering the finished product to the customer on time, but not too early.

The unique feature of this system is the use of time to measure the buffer size in the scheme of things to achieve right time delivery to the customers.

The major point to note here is that the buffer size was always computed as an approximation and not as an exact measure. And the buffer size used was a liberal buffer. Liberal means an additional safety factor was added to the required time for the usual buffers.

A traditional DBR system follows the framework denoted below where M3 is the CCR. This framework shows the planning methodology for the system (Fig. 1)

The point to note in these lists is that they are all schedules. Also, the CCR is the only resource to be scheduled. Rest of the resources run on a FIFO format. This approach is known as subordination of all non critical resources to the schedule of the constraint resource. The lean world knows this as the pacemaker process. The uniqueness of the TOC approach is in ensuring that the raw materials are not introduced too early into the system and have a specific date of release into the system depending on the buffer size. This is evident on close inspection of the dates mentioned on the various schedules in the example.

This framework is kept under control by monitoring the buffer penetration of the work. Each buffer was divided into three zones. The lowest zone is also known as the

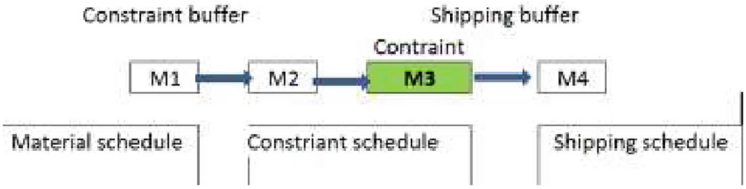


Fig. 1. Two types of buffers in DBR to ensure the flow (2000) [8]) (Source: Schragenheim. & Dettmer

red zone which denotes the need to expedite the process to meet customer demand date. The buffer itself is a rough estimate of the lead time plus the safety time needed to move the product from one stage to the finishing stages of the production process. And this time buffer usually deals with total time needed for a batch to be produced rather than a per product time as is the case in lean approaches. And there is no co mputation rule for the buffer size or CCR rope release schedule (Fig. 2).

Fig. 2. Buffer Zones to monitor for each order (Source: Schragenheim. & Dettmer (2000) [8])

2 Simplified DBR Principles

Simplified DBR (SDBR) was proposed by Schragenheim. & Dettmer (2000) [8]. While DBR was indeed a radical new method of approaching production situations, it still needed to find out a CCR within the manufacturing environment and had the burden of managing three buffers. In addition it was needed to schedule this resource and subordinate other resources to the working of this resource. When a system makes multiple products, each having a different CCR, this method tends to become that much more complex to deal with. Also a system makes money (Throughput) only by satisfying customer demand. While a single CCR decides the system's capacity, using that to manage the system can make operations complex, especially when multiple products are being made with multiple resources. The SDBR approach proposes a slight variant to dealing with this situation. Here only one buffer is used, the shipping buffer. This was proposed as most production systems have the market as the bottleneck limiting the earning capacity of the system. In addition to using the shipping buffer to track individual orders for tardiness, system loading also needs to be monitored. While system capacity still needs to note the capacity of the CCR, no organization will load the system beyond this capacity. So the market decides the load on the system. This paper does not consider

system loading as the simulation does not consider multiple periods of production. In most cases the market demand is less than the CCR. Hence it is recommended to use the market as the constraint and subordinate all the resources loadings to this demand. This framework is shown below Fig. 3.

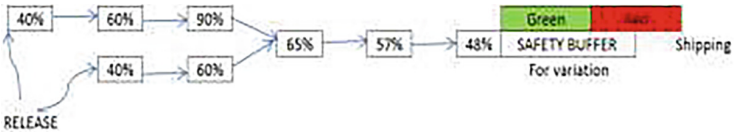


Fig. 3. SDBR framework (Source: Schragenheim. & Dettmer (2000) [8])

To control the process, SDBR has introduced the concept of planned load. Planned load is the total time (in hours) required of the CCR to complete all work that has been formally released into the system, but which has not yet been processed at the CCR Schragenheim & Dettmer (2000) [2]. The diagram shows the loading percent of each resource. It is evident that the CCR should have the highest loading among the resources. One extra point added in the SDBR framework is that no resource (including the CCR) is to be loaded beyond the 90% mark. This will lead to plenty of disruptions which come with internal variances. Another point of variance is the focus on the red zone more than using all three zones in the time buffers being used. The buffer used in both traditional DBR and SDBR is a time buffer, a sort of estimated lead time that is promised to the customer for delivering a batch of products accepted to manufacture. This time is not an exact estimate and carries a liberal buffer with it. If the buffer is too short, it will lead to defaulting on due dates, leading to a loss of reputation. On the other hand if the buffer is too large, there will be lack of utilization of the resources in the system and also a loss of reputation when compared to other players in the market. There is no strict rule for deciding the buffer size, except a thumb rule which states that 75% of the delivery period can be a good starting point. The buffer is the deciding factor for the release of the raw materials into the manufacturing system. If the buffer is too big, then materials would be released too early into the system and if the buffer is too small, the material will get in too late, leading to a starving of the resources.

Another variable of interest is splitting the order into smaller parts before initiating the manufacturing process. Meaning: an order of 100 pieces when split into 2, will mean that two batches of 50 items are made in separate batches. The higher the number of splits, the smaller will be the injection of material into the system for each production order release. This has nothing to do with the final delivery dates to the customer and only is a convenience factor in managing production.

The control of the schedules is achieved using two actions

1. Monitoring the Red line crossings of each batch
2. Monitoring the capacity loading of the CCR

Each buffer is the estimated time of delivery of the order to the customer. This is represented by the green zone, where the order would get delivered in the normal course of action. But when Murphy's Law starts applying, delays happen and it becomes

essential to add an additional padding to this delivery time [8, 9]. While there is no standard method prescribed, drawing from TOC-CCPM literature, we can add 50% of the estimated delivery time as the additional buffer time. This zone is referred to as the Red zone or the red line crossing in SDBR literature. When an order crosses over into the red zone it indicates that the order has been delayed and needs urgent expediting. Expediting can be done in any manner, but usually it is in the form of preferential processing access to the CCR and use of other expediting methods like overtime, outsourcing, additional working hours, etc. It is best not to have any red work orders in the systems. More than 5 can lead to a total collapse in delivery schedules.

By monitoring the loading level on the CCR. If this value is less than 90% then the system is stable. On the other hand if this resource loading exceeds 100%, it indicates an overload and an imminent resulting red line crossing of one or more orders leading to default on customer due date. Once the order has finished processing on the CCR, the load value of the processed product is removed from the CCR loading, freeing up the system for accepting future orders even though the product might still be in later stages of production past the CCR. Any excess of Red line crossings need to be expedited. It is a generally accepted rule that more than five red crossings lead to a collapse of the manufacturing system.

3 Research Question

It is a commonly accepted principle that buffers are required to safeguard against any contingencies. The greater the contingency, the larger is the buffer required. Even in environments where the processing times are certain, the lead tiles are stable and the customer demand does not change, these buffers play an important role in meeting customer delivery dates. It is therefore a requirement to have adequate buffer sizes. Hence we propose the hypothesis

H1: Longer buffers lead to better protection of customer deliveries

Every customer demand can be processed as one single batch or multiple batches of smaller quantities. Multiple batches increase the complexities of scheduling, but has lesser material injected into the system for each sub order released. This also loads the resources to a lesser extent. The bigger batches take a longer time to get processed on the CCR. Smaller batches leave the CCR faster, thus releasing capacity on the CCR more frequently. The loading on the CCR too is much lesser for each sub batch being released into the system. This means lower work in process stocks within the system, meaning less clogging of the system. However, larger sub batches imply a larger number of setups and can create new constraints in the system, especially for longer set up resources. However within normal restrictions, splitting batches into two or more sub batches leads to a better utilization of the resources in the manufacturing system which has been a long known fact in the lean manufacturing world, where they call it smaller batch size loading. Hence we hypothesize the following

H2: Higher splits (spread) in a customer delivery batch, improve the customer delivery performance even for smaller buffer sizes

4 Experimental Setup

This experiment uses Goldratt Simulator (GSIM) as mentioned in Goldratt & Cox (2003) [3] to generate the outcomes of the various experimental conditions being tested. The simulator has standard configurations of manufacturing facilities. This paper uses *Config. 310* which has a general configuration and five resources - blue, green, cyan, magenta and brown. Each has the following particulars

	Blue	Green	Cyan	Megenta	Brown
Set up time (mins)	15	120	60	30	0
Number of resources	1	2	2	2	1
Available processing time (hrs) including all resources	40	80	80	80	40

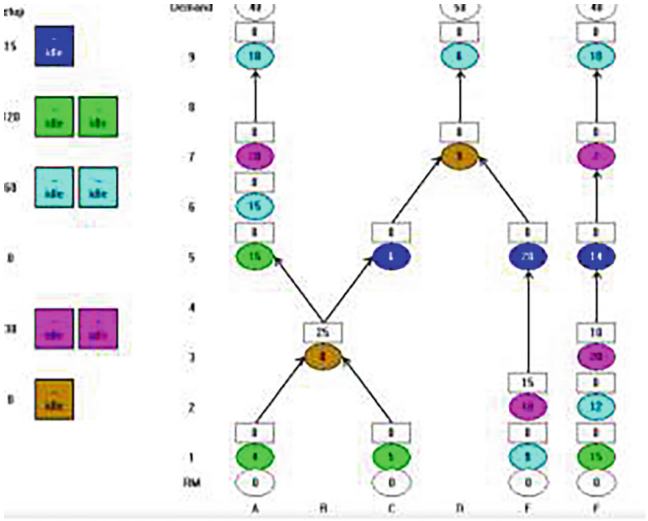


Fig. 4. A layout of Goldratt simulator game for production the TOC way (Courtesy: Eli Schragenheim, www.goldrattgroup.com/tocsimulator)

This is a paradise plant (Fig. 4). Every resource is reliable and never fails to deliver the load assigned to it, the quality output is perfect, vendors are reliable and efficient and always deliver on time, the machines never break down, customer order and delivery date is known. Added to that, the simulator has only one delivery deadline, namely one week lead time of 8 hours a day. Barring setup time, the resources can work all the remaining time. It is noticed that automating the game rarely yields results unless proper policies are framed.

The overheads of the plant are $11,000 for the week. The throughput for the products A, D and F are $180, $240 and $180 respectively. The order volumes for the products are 40, 50 and 40 respectively. These too are kept as a constant. Added to this, there is no

continuing option of simulating the game for the future weeks. In spite of the conditions being near ideal, most players who play the game using their cognitive genius rarely cross a net profit of $4,500 out of a maximum possible $5,450 (Goldratt & Cox, 2003) which indicates the need to have a clear policy driven approach to running the simulator to get best results.

In this backdrop, this paper uses different settings for the buffer size and the order split as follows

Buffer sizes: 10, 20, 25, 30, 40

Order spread: 1, 2, 3, 4

These are run as a full factorial design experiment. In addition the following points are used in the settings

1. On the CCR option, no constraint is mentioned as we are using SDBR approach
2. Below this option, we enable "Schedule common part operations"

Point 2 has been incorporated into the experiment as the simulator does not define and allocate specific materials for each order in resources which process similar materials for two different subsequent processes.

The split (spread) is entered in the option "Free products timing" and four values of splits are entered for all the three products A, D and F as an equal spread. Next, the rope signal is enabled for these settings, which allows the materials to be pulled into the system at appropriate times. The rope serves as an assignment timetable for the materials entering the system.

The game is then run in an automated fashion over five days and tested for throughput and profitability of the strategy. The resources are allocated where ever the load is found. Allocation of resources to a job is done in a random fashion. The buffer is not monitored, nor is the resource load monitored as there is only one order to deliver and that too at a fixed future date. The results are summarized and discussed in the section below.

5 Results and Hypothesis

The hypothesis is also tested using full factorial ANOVA. The descriptive results of this test too are tabulated (Table 1).

The Table 1 summarizes the results of the experiment. It is possible to spot a clear increase in profitability as one move downwards or towards the right in the table indicating that the hypothesis proposed is sufficiently strong and worth testing. This is also graphically demonstrated below (Fig. 5).

The generated data in the simulation run of GSIM under the given experimental conditions are tested using the full factorial ANOVA test. The results of the test are summarized in the Table 2.

Rope: The effect of the variable Rope is highly significant ($p<0.001$), indicating that changes in Rope contribute substantially to the variability in the profit generated.

Spread: Similarly, Spread is also a significant predictor ($p<0.001$), indicating that it plays a significant role in explaining the profit generated.

Rope * Spread: The interaction between Rope and Spread is significant ($p<0.001$), implying that the effect of Rope on the profit generated depends on the level of Spread.

Table 1. Summary of experiment outcomes

Average of Net profit	*Spread*			
Rope size	1	2	3	4
10	-6515	-2775	512.5	1552.5
20	-5120	2620	5450	5450
25	5450	4932.5	4990	5305
30	4875	5450	5450	5105
40	5450	5450	5450	5450

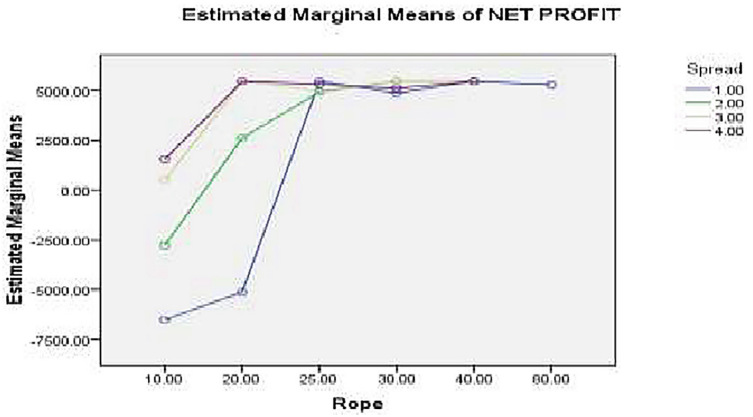


Fig. 5. Line graph showing the rope size and number of splits (spread) (MiniTab output)

The test indicates that three hypothesis are statistically significant at a probability of 0.01

The size of the rope has a significant impact on the profit generated in the simulation run (H1)

The spread has an impact on the impact of the rope size on the profit generated in the simulation run (H2)

And the test also indicates a third hypothesis as significant, though it was not proposed initially. Since this result is of immense interest to researchers, this is also being mentioned as a possible future research question for study.

H3: The spread by itself has a direct impact on the profit of the simulation run

Table 2. Factorial ANOVA test results (Dependent Variable: NET PROFIT)

Source	Type III Sum of Squares	df	Mean Square	F	Sig
Corrected Model	4.016E8	20	2.008E7	35.495	.000
Intercept	3.917E8	1	3.917E8	692.312	.000
Rope	**2.490E8**	**5**	**4.981E7**	**88.037**	**.000***
Spread	6.705E7	3	2.235E7	39.506	.000*
Rope * Sprea1.158E8d	1.158E8	**12**	**9650490.508**	**17.057**	**.000***
Error	9052412.500	16	565775.781		
Total	8.698E8	37			

* Statistically Significant at 1% significance level

(Note: R Squared: The model explains 97.8% of the variance in the dependent variable, and the adjusted R Squared value of 0.950 indicates that the model is robust, accounting for the number of predictors in the model)

6 Conclusion

This paper has demonstrated the use of simulation in testing out the impact of production planning policies on the flow of material through a production system. This research paper has considered the impact of two variables, the rope size and the split in order size.

This research paper has shown the statistical significance of two long-standing hypotheses used in the TOC circles for decades. It has conclusively established the significant impact of adequate buffer size on throughput and hence profit generated by a manufacturing system. Smaller buffer size leads to shorter response time which cause a lot of concern related to finishing the work on hand. There is a log of competition on resources as well just to complete the orders. Bigger buffers give sufficient time to finish the work undertaken.

It has also proved the statistically significant impact of the splitting of an order into smaller bits to improve performance of the system under smaller buffer size conditions. Shorter production runs allow us to work with smaller buffers to achieve the same production effectiveness. The splitting of orders significantly improves the profitability of the enterprise. Only when the split becomes too small, especially when set up times are a significant part of the production time, we tend to face difficulties in completing the work due to multiple stoppages and time waste within the system. But for these issues are very high level of order splits, the general benefits were smaller material requirement by the system and frequent small purchases. This allows conservation of cash and other resources if the system

Most experts however, are of the opinion that when buffer size is increased beyond a point [10] (Schragenheim et al, 2009), the performance would tend to suffer, especially

when there is a constraint of cash available to buy raw materials and one needs to generate throughput to finance future purchases of raw materials [9] (Schragenheim and Dettmer, 2000). This could be the topic of future research papers which could consider other software packages to work on other experimental variables like future week demand, demand acceptance, buffer monitoring, expediting actions recorded, etc. However, this paper has established simulation as a robust tool to test hypothesis related to TOC framework, which could otherwise be only intuitively arrived at by experience or by the instruction of someone with many years of experience in implementing TOC in the industry. But in either case, this kind of data or statistical testing would be impossible to conduct.

References

1. Aggarwal, S.C. (1985), "MRP, JIT, OPT, FMS?", Harvard Business Review, Vol. 63 No 5, pp. 8–16.
2. Corsini, R.R., Costa, A., Cannella, S., Framinan, J.M.: Analysing the impact of production control policies on the dynamics of a two-product supply chain with capacity constraints. Int. J. Prod. Res. **61**(6), 1913–1937 (2023)
3. Goldratt, E.M., Cox, J.: Production the TOC way. Great Barrington: North river press (2003)
4. Goldratt, E.M., Cox, J.: The goal: a process of ongoing improvement. Routledge (2016)
5. Johnson, A.: MRP? MRP II? OPT? CIM? FMS? JIT? Is any system letter- perfect? Manage. Rev. **75**(9), 22–27 (1986)
6. Koziol, D.S.: How the constraint theory improved a job shop operation. Management Accounting, vol. 69, No. 11, pp. 44–49 (1988)
7. McCleskey, J.A.: Forty years and still evolving: The theory of constraints. Am. J. Manage. *20*(3) (2020)
8. Orue, A., Lizarralde, A., Amorrotu, I., Apaolaza, U.: Theory of constraints case study in the make to order environment. Journal of Industrial Engineering and Management (JIEM) **14**(1), 72–85 (2021)
9. Orue, A., Lizarralde, A., Apaolaza, U.: The impact of protective capacity in implementing drum-buffer-rope methodology in make-to-order environments: an assessment by simulation. Dyna Manage. J. 2013 **10**(1) (2022)
10. Panizzolo, R.: Theory of constraints (TOC) production and manufacturing performance. Int. J. Ind. Eng. Manage. **7**(1), 15–23 (2016)
11. Schragenheim, E., Dettmer, H.W.: Manufacturing at warp speed: Optimizing supply chain financial performance. CRC Press (2000)
12. Schragenheim, E., Dettmer, H.W.: Simplified drum-buffer-rope: A whole system approach to high velocity manufacturing (2000)
13. Schragenheim, E., Dettmer, H.W., Patterson, J.W.: Supply chain management at warp speed: Integrating the system from end to end, Auerbach Publications (2009)
14. Wu, K., Zheng, M., Shen, Y.: A generalization of the Theory of Constraints: Choosing the optimal improvement option with consideration of variability and costs. IISE Transactions **52**(3), 276–287 (2020)
15. www.goldrattgroup.com/tocsimulator

Effect of Technology Acceptance Among Employees in the Second Half of Their Working Life

Firdaus Bashir(✉)

PSG Institute of Management, PSG College of Technology, Coimbatore 641004, India
firdaus@psgim.ac.in

Abstract. Integration of technology into the workspace is a continuous process which has a major impact in the manner of functioning. It is essential to make the employees ready in accepting and integrating technology in their work. The present study explores the effect of socio-demographic factors like age, job designation, educational qualification and years of experience of employees who are in their second half of working life on the acceptance of technology integrated in the organisation. Four dimensions of acceptance of technology was considered namely usefulness of technology integrated at work, user-friendliness of new technology, availability of training and organisational support. The Technology Acceptance Model (TAM), was applied as the grounded theory for the present study. Survey method was integrated where 250 structured questionnaires were distributed to the supervisors and clerical employees of three manufacturing organisations, out of which 225 were further processed for percentage analysis, chi-square and the correlation coefficient test. It was observed that organisations should adopt technology which is useful as well as user friendly. Providing sufficient training and organisational support will make the acceptance effective. Management can use the insights of the study to adopt use-friendly technology, provide sufficient training programs for employees of different age groups, educational backgrounds, years of experience and designation to ensure acceptancy.

Keywords: Socio-Demographic factors · Technology Acceptance · Supervisor · Clerks · Employees in the second half of working life

1 Introduction

The advent of technology has tremendously transformed the nature of the workplace. It redefines jobs and exerts new occupational demands on employees [1]. Resisting this inevitable technological change, employee skill sets become obsolete and outdated, finding themselves unemployable. Organisations integrating technology are increasingly training their workforce to avoid any mismatch between the skills required and the skills available [2] for both younger and older employees [3]. Past literature highlights that older employees tend to evaluate their digital competencies more negatively than younger employees [4]. Such a gap affects productivity and increases older employees'

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 348–362, 2025.
https://doi.org/10.1007/978-3-031-86305-9_24

vulnerability to unemployment. The present study considers employees who are in the second half of their working life which means that they have crossed their midpoint and are presently in their later stages of professional life, graciously moving towards retirement. Employees in this age category are on the rise, which is a crucial turning point in their careers. This is because they must deal with multiple issues, including health concerns and changes in professional motives and perspectives. The second half of working life usually commences around the age of 40 years, which is pivotal for planning and developing careers ahead. In the discourse surrounding the socio-demographics of employees, due attention has been allotted to "younger" and "older employees", frequently overlooking a considerable segment of the workforce under the category "middle-life employees". Young employees draw attention because they pose a threat to the acceptance of a newly adopted organisational policy [5]. In contrast, older employees may face doubt about their ability to keep up with the new job demands as they age. Employees in their mid-career are seldom addressed in terms of socio-demographic characteristics, especially in the realm of technology integration.

1.1 Context of the Study

Organisations in the manufacturing sector enhance their productivity and competitiveness by adopting various technologies [6]. They are constantly urged to retain customers, survive confrontations, produce cost-effective products, consider consumers demand, manufacture customised products that are economical and maintain greater quality. Adopting and accepting technology has benefited these organisations by enhancing flexibility, efficacy and accelerating the production process. To obtain the best output from technology adoption, all employees, including those in the second half of their working lives must effectively adopt it. It is important to understand the socio-demographic factors of employees accepting technology. These socio-demographic factors not only affect individual behaviours [7] but also have implications for organizational strategies [8] that foster equitable technology adoption. A deep understanding of socio-demographic factors such as age, educational qualification, job designation, and work experience will enable organisations to design targeted interventions such as training or organisational support that could address the specific needs of employees. The study exclusively focuses on supervisors and clerks of the selected manufacturing organisation. Supervisor is the one who exhibits supervisory behaviour to the subordinates shaping their work environment. They provide a link between employees and management systems [9]. They have major responsibilities such as overseeing operations, troubleshooting issues, allocating resources carefully to prevent bottlenecks, ensuring a smooth flow of performance [10], and most importantly monitoring how well technology is accepted [11]. Clerks who hold clerical job designations in manufacturing organisations and engages in administrative and support tasks are the other segments of employees focused in the study. They are responsible for maintaining various financial records [12], ensuring the continuity of supply chain, managing resources [13], dispatching finished products on time [14] and maintaining compliance with legal requirements [15]. If they do not adapt to technology, the consequences for clerical staff could be severe as they could even be displaced from their jobs. The available literature on clerical employees, especially in their midlife careers, is scarce. Therefore, further research on this topic is important.

1.2 Theoretical Underpinning

The Technology Acceptance Model (TAM) [16] was integrated as the grounded theory for the present study. It holds on two components, namely Perceived Ease of Use (PEOU) and Perceived Usefulness (PU)—the former being perceived ease of use which means that the technology is perceived to be simple to use, and the latter being perceived usefulness, which means that the technology when put to use can improve productivity; these two aspects can either develop a positive or negative attitude towards the use of technology. TAM has been refined over time to other models such as TAM2, TAM3, the Unified Theory of Acceptance and Use of Technology (UTAUT). This theory is foundational to comprehending user acceptance of technology in various contexts, including the influence of socio-demographic factors on technology integration. It has been extensively validated across different domains and in an employee context. Through this theory, different social demographic factors such as age, gender, education level, years of experience, and impact on acceptance of new technology can be explained. Previous literature has studied the influence of demographic factors on technology acceptance using this theory. Older employees perceived the integration of new technology as difficult compared to younger employees. Correspondingly, educational qualifications and years of experience can shape a positive perception of the usefulness of an integrated technology [17]. When technology is customised to the diverse needs of demographic groups, its acceptance and implementation improves. Therefore, the manufacturing industry too can integrate technology to suit various employee needs and ensure smooth acceptance.

2 Review of Literature

Socio-demographic factors are significant in research for understanding the interplay between the characteristics of a person and numerous outcomes. Age affects various outcomes such as cognitive development, health status and life experiences [18]. Besides age, a proxy for experience, is a stronger predictor of affective commitment than are other demographic variables [19]. Years of experience is another factor which denotes the duration of an individual's exposure to professional realms, building a knowledge base and skill sets [20]. Several levels of educational qualifications have numerous outcomes. It is observed that when formal education is low in employees, their participation in training programs tends to diminish [21]. This could be due to a lack of motivation, or encouragement from management or not having enough awareness about the benefit of attending training programs. Lifelong learning is required to maintain productivity, particularly with employees in the second half of their working life. This is because of the psychological strain suffered when confronted with the difficulty in learning a new technology integrated into their work. To avoid this, organisations should invest in targeted training, equipping them with essential skills to accept and adapt a new technology introduced [22]. It is pertinent for both employees and employers to consider all the career phase that requires skill enhancement. Socio-demographic factors are also found to negatively impact preferences and perceptions of individuals. Therefore, it is interesting to note that these factors have diverse implications for research [23]. Especially in studies where employees need to inculcate new work-related capacities [24]. Introduction of a new technology is associated with an increased level of stress,

as it demands multitasking in supervisory roles [25]. Besides lack of adaptation can cut down jobs too, as lot of clerical work has been substituted by technology [26]. Hence, updating skill hence has become mandatory to stay relevant and current. Supervisors are vital part of the workforce and are expected to be vivid and innovative. They have to be agile in response to the growing demands of complex challenges including technical revolutions, inclusiveness and changing responsibilities.

2.1 Research Gap and Research Questions

Despite the growing body of research on the acceptance of technology, an explicit gap persists in understanding how employees in clerical and supervisory cadres specifically those in the second half of their career reconcile their potential with learning a new technology. They are an extremely important segment of the workforce and are growing in number. The present study helps to mitigate this gap by understanding how they navigate technological transitions. Past literature has focused on younger and older employees, with less emphasis on the socio-demographic factors of mid-career employees, who may face distinct challenges. Exploring the link between social demographic factors and the adoption of technology particularly among these employees is drawing the attention of researchers, organisations and policymakers. For an organisation to enjoy the benefits of technological advancement employees should use it for a wide variety of tasks in hand. This study aims to understand the socio-demographic factors of clerks and supervisors and their significance in the acceptance of technology. Hence the following research questions were developed for the study (RQs):

- RQ1: To what extent does socio-demographic factors such as age, job designation, educational qualification, and years of experience of supervisors and clerks have an impact on the usefulness of a new technology integrated?
- RQ2: How does the socio-demographic factors of these employees who are in their second half of working life effect the user-friendliness of the new technology integrated?
- RQ3: Do these factors have an impact on the training available to accept the technology in the organisation?
- RQ4: Can socio-demographic factors of these employees influence the support available in the organisation in accepting the new technology adopted?

3 Research Methodology, Measures and Tools

Before administering the questionnaire, a cover letter explaining the purpose of the study and guaranteeing confidentiality was offered to the management of the manufacturing organisations. In the questionnaire four items were identified and evaluated including 1) Usefulness of technology integrated, 2) User-friendliness of new technology, 3) Availability of training on technology integrated and 4) Organisational Support to measure the technology acceptance by supervisor and clerks who are in their second half of the working life. To measure the usefulness of technology integrated items like a) The new technology integrated enables in completing the task quickly, b) It helps in improving the quality of work in hand, c) The new technology enhances the productivity of an

employee, d) It simplifies complex work and e) It adds value to the designation the employee holds was adapted [16].

The five items like a) The new technology integrated can be learned quickly, b) It has user-friendly features c) It is easy to understand, d) Employees can develop proficiency after consistent use of the technology and e) Employees feel comfortable using it, were used to gauge the variable user-friendliness of new technology adopted in the organization [27]. To analyse the extent to which an employee receives training on the technology adopted, five items were a) Hands-on training opportunity are provided to the employee by the organisation, b) Adequate trainings are provided, c) The training provided is useful, d) The training provided is informative and e) The training sessions delivered enables an employee to adopt the new technology with ease [28]. The degree to which the organisation supports the use of technology was measured by items integrated from including a) The organisation provided necessary resources required to integrate the new technology, b) Technical support is available for implementation of the new technology, c) The organisation invests in maintaining the technology, d) The technology is updated when required and e) The organisation encourages employees to execute the new technology [29].

The present research concentrates on supervisors and clerks who are in their second half of working life from three manufacturing organisations located in Coimbatore, which is also known as the Manchester of South India. The selected manufacturing organisations were rapidly adopting technology in different domains and provided digital training to its employees. The research data was gathered using structured questionnaires framed using English as well as Tamil language. Tamil is the state language where the organisations are located and all the employees were highly familiar with it. The researcher made herself available to explain any quires from the employees if required. Firstly, a pilot study was conducted among randomly selected group of 30, of which 15 were supervisors and the other 15 were clerks who were excluded while collecting the final data. The questionnaire was distributed to 250 respondents, of which 225 questionnaires were further used in the statistical tests conducted and the rest which was found incomplete were excluded from the study. Statistical Package for Social Sciences (SPSS) version 28 was used to analyse the data. Judgmental sampling was the sampling method used for the study, which allowed the researcher to select supervisors and clerical employees who are in the second half of their working life as respondents to provide valuable insights. This sampling method enabled the researcher to identify the targeted respondents by gathering expert opinions from various department heads. The researcher followed certain predetermined criteria in selecting the respondents. An employee who was a supervisor or a clerk and had at least three years of work experience in the same organisation were chosen as a respondent for the study. Besides, they should have a minimum educational qualification of a diploma. The final data was entered in to the SPSS software, it was checked for missing values or errors, later the categorical variables were assigned numerical codes and organised in to a data frame. A percentage analysis was used to describe the sociodemographic factors and years of experience and later mean score of various dimensions of technology acceptance was calculated. The chi-square test, a non-parametric test, was applied to measure the association between categorical variables such as age, job designation, educational qualification, and years

of experience with the technology acceptance of employees P-values measured to be less than or equal to 0.05 were considered statistically significant. The correlation coefficient was measured to evaluate the association between two variables and the direction of the relationship. By combining these two statistical analyses, the researcher developed a nuanced understanding of various socio-demographic factors and its significance to technology acceptance.

4 Data Analysis and Results

A percentage analysis was conducted on the socio-demographic factors selected for the study. Table 1 illustrates the percentage distribution of employees' age groups in the second half of their working lives, Table 2 represents the percentage distribution of job designation of these employees, Table 3 illustrates the percentage distribution of educational qualification, and Table 4 depicts the percentage distribution of the years of experience gained.

Table 1. Percentage distribution of the age group of employees.

Sl. No.	Age	Number of respondents	Percentage
1	40 years to 44 years	25	11.11
2	45 years to 49 years	56	24.89
3	50 years to 54 years	90	40.00
4	Above 55 years	54	24.00
	Total	225	100

Source: Survey data

Table 1 depicted that the 25 employees, representing 11.11% of the total sample belonged to the age category of 40–44 years, 56 employees representing 24.89% of the total sample belonged to 45–49 years, 90 employees representing 40.00 percentage of the total sample belonged to 50–54 years, and 54 employees representing 24.00% of the total sample were either 55 years or above. Figure 1 illustrates the graphical representation of the data.

Table 2 displayed that 130 employees amounting to 57.78% were supervisors and the remaining 95 employees amounting to 42.22% were clerical employees.

Table 3 indicated that 63 employees amounting to 28.00% of the sample size were postgraduates, 101 employees estimating to 44.89% were graduates, and the remaining 61 employees forming 27.11% were diploma holders. The majority of respondents 44.89% were graduates.

Table 4 represented the years of experience of 225 employees. Ninety employees amounted to 40% had work experience ranging from 3–10 years. 105 employees estimating to 46.6% of the total sample had 11 years to 18 years of experience and they were in majority, 30 employees measuring to 13.3% had 19–26 years of experience and Table 5 displays the mean scores for various dimensions of technology acceptance in the

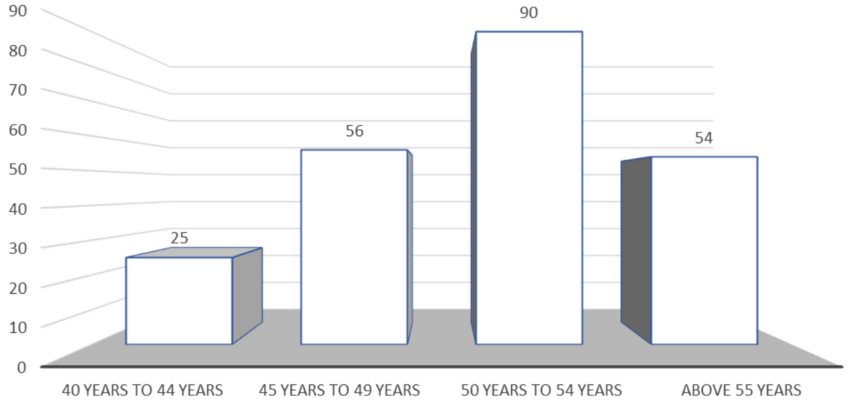


Fig. 1. Bar graph indicating number of respondents in different age groups

Table 2. Job designation of the employees in their second half of working life.

Sl. No.	Job designation	Number of respondents	Percentage
1	Supervisor	130	57.78
2	Clerical Employees	95	42.22
	Total	225	100

Table 3. Educational qualification of the employees who are in their second half of working life

Sl. No.	Educational qualification	Number of respondents	Percentage
1	Post Graduate	63	28.00
2	Graduate	101	44.89
3	Diploma	61	27.11
	Total	225	100

Source: Survey data

manufacturing sector. The mean score for user-friendliness of new technology integrated was (3.981) which means that employees found the features of technology integrated in the organisation to be user-friendly. The mean scores of usefulness of technology integrated at work (3.761), availability of training on technology adopters (3.018), and Organisational support had a means score of (2.982).

Chi-Square Analysis. The chi-square test was used to determine the socio-demographic factors and the impact on the acceptance of new technology among employees in the second half of their working life.

In Table 6, the chi-square value of 23.783, with a p-value of 0.001, indicates that the impact of the perception of usefulness of technology integrated across various age groups

Table 4. Employee's years of experience who are in their second half of working life

Sl. No.	Years of experience	Number of respondents	Percentage
1	3–10 years	90	40
2	11 years 18 years	105	46.6
3	19–26 years	30	13.3
	Total	225	100

Source: Survey data

Table 5. The mean score of the dimensions of technology acceptance

Dimensions	Mean Score
Usefulness of technology integrated at work	3.761
User-friendliness of new technology	3.981
Availability of training on technology integrated	3.018
Organisational support	2.982

Table 6. Socio-demographic factors and usefulness of the technology integrated.

Sl. No.	Chi square value	Calculated Value	P – Value	Result
1	Age Group	23.783	0.001	Significant
2	Job designation	21.774	0.001	Significant
3	Educational qualification	17.890	0.001	Significant
4	Years of experience	20.781	0.001	Significant

Source: Computed data, * 5% significant level

are statistically significant. This suggests that employee age plays a vital role in how individuals perceive technology usefulness. A chi-square value of 21.774, with a p-value of 0.001, indicates that job designation significantly impacts the usefulness of technology integration. This implies that the job designation an individual holds, for the present study which is supervisor and clerical category, was considered to have an influence on the usefulness of new technology integrated. A chi-square value of 17.890, with a p-value of 0.001, indicates that the level of education significantly impacts technology acceptance. This implies that the educational qualification of being a diploma holder, graduate, or postgraduate influences the likelihood of the usefulness of the new technology. The chi-square value of 20.781, with a p-value of 0.001 indicates that years of experience significantly impact the usefulness of technology acceptance.

Table 7 displays that years of experience with a chi-square value of 1.693 with a p-value of 0.535 does not significantly influence the user-friendliness of the integrated technology at a 5% significance level. Whereas age group with the chi-square value of

Table 7. Socio-demographic factors and user-friendliness of the technology.

Sl. No.	Chi square value	Calculated Value	P – Value	Result
1	Age Group	17.904	0.001	Significant
2	Job designation	15.561	0.001	Significant
3	Educational qualification	18.294,	0.001	Significant
4	Years of experience	1.693	0.535	Not-Significant

Source: Computed data, * 5% significant level

17.904 with a p-value of 0.001, chi-square value of job designation being 15.561 with a p-value of 0.001 and the chi-square value of educational qualification being 18.294 with a p-value of 0.001was found significant. Employees in the second half of their professional careers are concerned about the user-friendly features of the integrated technology. Technology can perform both simple and highly complex tasks. Different job designations such as supervisors or clerks have diverse job responsibilities. For example, the former needs to maintain employee records of progress along with mid-managerial responsibilities. Clerical employee's prerequisites may also differ, but both segment of employees prefer to accept user-friendly technology. Furthermore, employees with higher educational qualifications might find the integration of technology easier to understand than those with lower educational qualifications requiring more training.

Table 8. Socio-demographic factors and availability of training on technology integrated

Sl. No.	Chi square value	Calculated Value	P – Value	Result
1	Age Group	1.783	0.173	Not Significant
2	Job designation	17.632	0.001	Significant
3	Educational qualification	18.431	0.001	Significant
4	Years of experience	1.653	0.510	Not Significant

Source: Computed data, **** 5% significant level***

Table 8 represents the socio-demographic factors and availability of training. Years of experience 1.653 does not significantly influence the training availability at a 5% significant level. Age group with the chi-square value of 1.783 with a p-value of 0.173, too were found not significant. Chi-square value of job designation being 17.632 with a p-value of 0.001 and the chi-square value of educational qualification being 18.431 with a p-value of 0.001was found significant. The training provided within the organisation was made uniformly available to all the employees regardless of their age. Significantly, age seemed not a significant factor. Employees with different job designation require hands on training in relation to their work context. A technology integrated might vary in features, so do the training prerequisites by clerical employees as well as supervisors. Targeted training aligned with the diverse requirements of employees encourages acceptance of technology. Employees with lower educational qualifications when compared

to those with higher education might require more comprehensive training to enabling them implement the technology in their daily work. This disparity in learning needs to be addressed by the organisation for better acceptance.

Table 9. Socio-Demographic Factors and Organisational Support

Sl. No.	Chi square value	Calculated Value	P – Value	Result
1	Age Group	20.683	0.001	Significant
2	Job designation	12.367	0.003	Significant
3	Educational qualification	23.084	0.002	Significant
4	Years of experience	22.673	0.002	Significant

Source: Computed data, * 5% significant level

Table 9 indicates the socio-demographic factors and organisational support received in accepting the technology. Age group with a chi-square value of 20.683 with a p-value of 0.001, chi-square value of job designation at 12.367 with a p-value of 0.001, the chi-square value of educational qualification at 23.084 with a p-value of 0.001, and years of experience with the chi-square value of 22.673 with a p-value of 0.002 were found to be significant at a 5% significance level. Organisation supports employees by providing them with technical training and investing in maintaining and updating technology. Technology can be a double-edged sword for employees, if they do not have enough resources to learn and apply, the outcome would not be as expected. Therefore, management in the organisation should be very careful before introducing a new technology and provide employees with sufficient resources and training to acquire the required skills.

Correlation Co-efficient. The correlation tool was employed to discover the relationship between the four dimensions adapted for technology acceptance among employees in their second half of their working life.

According to Table 10, the usefulness of integrated technology has a strong positive association with the user-friendliness of new technology integrated at a 1% significance level; the correlation coefficient ($r = 0.913$) which is closer to 1. This signifies that the user-friendliness of integrated technology enhances its perceived usefulness. The Usefulness of technology integrated at work revealed a strong correlation coefficient with organisational support ($r = 0.814$) at a 5% significance level. When an organisation provides sufficient training and resources to employees, they will become confident enough to accept the new integrated technology. At the 5% significance level, the user-friendliness of the new technology integrated has a favourable relationship with the availability of training on technology integrated ($r = 0.782$). Training should be informative and useful, enabling them to handle their work with ease. At the 1% significance level, the user-friendliness of the new integrated technology has a positive relationship with organisational support ($r = 0.756$). The availability of training was positively associated with organizational support at a 5% significance level ($r = 0.673$). Organisational support is a crucial factor in the implementation and execution of technology. It should

Table 10. Correlation co-efficient and technology acceptance among employees

	Usefulness of the technology integrated	User-friendliness of the new technology	Availability of training	Organisational support
Usefulness of the technology integrated	**1**			
User-friendliness of the new technology	**0.913****	**1**		
Availability of training	0.641	**0.782***	**1**	
Organisational support	**0.814***	**0.756****	**0.673***	**1**

provide the necessary resources, technical support, timely maintenance and upgradation to receive a higher rate of acceptance from employees.

5 Discussion

Embracing technology is the most prominent part of contemporary society enabling the smooth running of day-to-day business activities [30]. Technological advancement represents an opportunity for accelerated social and economic growth [31]. For an organisation to enjoy these benefits, the willingness of all the employees to accept the technology is paramount [32]. Socio-demographic factors, such as age, income, experience and educational qualification are associated with how an individual gets acquainted with technology. Age has been found to have a positive relationship with technology acceptance among college teachers [33], and work experience and educational qualification have influenced e-banking adoption in India [34]. Likewise, the results of the present research demonstrate that the socio-demographic factors have a significant impact on the adoption of technology. Cognitive and psychological changes are among various factors that make older employees reluctant to accept technological advancement [35] because ease and comfort become a priority for them [36]. Technology without user-friendly features may require employees to spend excessive time exploring it. Therefore, focusing on adopting a time-saving and user-friendly one enables an easy acceptance among employees who are in the second half of their working life. Moreover, when it seems useful and relevant to a job in hand the acceptance becomes smoother. All the select socio-demographic factors except work experience were found to have a significant relationship with the user-friendliness of the technology adopted. The TAM model holds that perceived usefulness and ease of use are beliefs that influence the use of technology [16].

Training delivered to employees can be tailored to address explicit needs and decrease the skill gap required for using the new technology. Unfamiliarity with a new integrated

digital tool within the organisation manifests fear, leading to low confidence and self-esteem that will trigger a feeling of threat, forcing employees to return to their old fashion of work. Training can boost confidence in them which is the key to learning a new technical skill [37]. Accepting a technology will enable the supervisor to accomplish work effectively and engage in participation as and when subordinates report to him for clarifications about the new technology. They are responsible for confirming on employees updating their skill-sets for which they have to apprise their own competencies. If clerical employees do not keep themselves abreast, they might lose jobs as technology is capable of performing most clerical jobs by itself [38]. They can climb the career ladder only when adapting to newer technology. New technology adopted in an organisation will enable them to reinvent various ways of doing a job. These employees in the second half of their career life have to continue being productive and upscale according to the new demands of the job. For which adequate training, resources, and technical support should be ensured by the organisation.

Active ageing has attracted global attention. Employees who age gracefully gain rich experiences that can contribute to organizational growth, contrary to the traditional belief that they are unproductive [39]. However, they have to equally get accustomed to technological advancement when organisations integrate new digital tools [40]. This transition has not only changed the work system, but also the employee identity [41], relationships at work [42], and work content [43]. The acquired technical knowledge may expire within two to three years therefore employees are necessitated to upskill themselves periodically. Management should take due care in up-skilling the employees while integrating workplace technologies [44], especially those who are in the second half of their careers by providing the required resources through training and development to stay current [45].

5.1 Implication of the Study

The study implies that employees in different phases of their careers, especially those in the second half of their working life, should be sustained as contributors to the organisation as well as the economy at large [46]. These employees often were overlooked especially in technology acceptance domain. Therefore, policy makers should make policies that encourage continuous learning and skill development among them. The study contributes to the literature by offering a holistic understanding of different socio-demographic factors and their associations with various attributes of technology acceptance. This is also a niche area where the literature availability is comparatively scarce. The present study adds to the existing body of knowledge by integrating TAM model into technology acceptance. Top management and policymakers should ensure the usefulness and ease of use before they decide to adopt any technology. Furthermore, organisations should customise training programs for employees belonging to different age groups, educational backgrounds and designations. Employees in their second half of working life are often characterised by challenges of continuous upskilling and adaptation to changes, therefore, focusing on them is crucial for enhanced employee productivity and organisational growth. The study implies appropriate sufficient training as well as organisational support [47] which will help management minimise the disparity that exists between the skills required to navigate a new technology effectively by the employees.

5.2 Limitations and Future Research

This study had certain limitations that must be acknowledged. The data were collected from the manufacturing sector therefore, the results may not reflect the reality of other industrial sectors. The generalisability of the results is a limiting factor since supervisors and clerks were the only employees focused in the study. In future researchers can concentrate on employees holding job designation and extend the study to other industrial domains like health sector or tourism. The study concentrated on organisations that are technically advancing at a rapid phase. The ones which were not in this phase were excluded. Conducting longitudinal studies to examine whether sociodemographic factors change over time would provide interesting insights. Researchers can also conduct qualitative research leading to thought-provoking insights. Apart from the sociodemographic factors tested, future studies can evaluate other factors like gender or income. It can also concentrate on different geographic regions to understand the degree of technology acceptance among employees in their second half of working life.

References

1. Frey, C.B., Osborne, M.A.: The future of employment: how susceptible are jobs to computerisation? Technol. Forecast. Soc. Chang. **114**, 254–280 (2017)
2. Brunello, G., Rückert, D., Weiss, C., Wruuck, P.: Advanced digital technologies and investment in employee training: complements or substitutes? J. Technol. Manage. (2023)
3. Lloyd, C., Payne, J.: Digital skills in context: working with robots in lower-skilled jobs. Econ. Ind. Democr. **44**(4), 1084–1104 (2023)
4. Anderson, M., Vogels, E.A., Perrin, A., Rainie, L.: Connection, creativity and drama: teen life on social media in 2022. Pew Research Center (2022)
5. Egdell, V., Maclean, G., Raeside, R., Chen, T.: Age management in the workplace: manager and older worker accounts of policy and practice. Ageing Soc. **40**(4), 784–804 (2020)
6. Gillani, F., Chatha, K.A., Jajja, M.S.S., Farooq, S.: Implementation of digital manufacturing technologies: antecedents and consequences. Int. J. Prod. Econ. **229**, 107748 (2020)
7. Huijts, N.M.A., Molin, E.J.E., Steg, L.: Psychological factors influencing sustainable energy technology acceptance: a review-based comprehensive framework. Renew. Sustain. Energy Rev. **16**(1), 525–531 (2012)
8. Deressa, B., et al.: The relationship between organizational commitment and organizational justice among healthcare workers in Ethiopian Jimma Zone public health facilities. J. Healthc. Leadersh. **14**, 5–16 (2022)
9. Rahim, N.B., Osman, I., Arumugam, P.V.: Linking work-life balance and employee well-being: do supervisor support and family support moderate the relationship? Int. J. Bus. Soc. **21**(2), 588–606 (2020)
10. Singh, V., Sharma, P.: Operational efficiency and workforce management in manufacturing. Manuf. Syst. J. **45**(4), 112–129 (2021)
11. Sinambela, E.A., Mardikaningsih, R., Arifin, S., Ayu, H.D.: Development of self-competence and supervision to achieve professionalism. J. Islam. Econ. Perspect. **1**(2), 33–42 (2020)
12. Lee, J., Wang, H.: Cost management through administrative functions in manufacturing. Asian Ind. Rev. **20**(3), 65–78 (2021)
13. Chand, A., Patil, R.: The importance of clerical roles in industrial production. Global Manuf. Rev. **12**(1), 89–103 (2020)
14. Mukherjee, S., Verma, K.: Operational coordination and the role of clerks in manufacturing. Ind. Oper. Rev. **14**(2), 47–58 (2021)

15. Gupta, P., Roy, A.: Compliance and reporting in manufacturing: the critical role of clerks. Corp. Comp. J. **15**(2), 245–261 (2022)
16. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User acceptance of computer technology: a comparison of two theoretical models. Manage. Sci. **35**(8), 982–1003 (1989)
17. Igbaria, M., Zinatelli, N., Cragg, P.B., Cavaye, A.L.M.: Personal computing acceptance factors in small firms: a structural equation model. MIS Q. **21**(3), 279–305 (1997)
18. Burger, R.J., et al.: Gestational age and socio-demographic factors associated with school performance at the age of 12 years: a population-based study. Paediatr. Perinat. Epidemiol. **37**(7), 643–651 (2023)
19. Mustafa, S., Harding, J., Wall, C., Crowther, C.: Sociodemographic factors associated with adherence to dietary guidelines in women with gestational diabetes: a cohort study. Nutrients **13**(6), 1884 (2021)
20. Giffords, E.D.: An examination of organizational commitment and professional commitment and the relationship to work environment, demographic and organizational factors. J. Soc. Work. **9**(4), 386–404 (2009)
21. Cedefop: Skills, qualifications, and jobs in the EU: The making of a perfect match? Evidence from Cedefop's European skills and jobs survey. Publications Office of the European Union (2016)
22. Obisi, C.: Employee training and development in Nigerian organizations: some observations and agenda for research. Aust. J. Bus. Manage. Res. **1**(9), 82 (2011)
23. Lyons, S.T., Ng, E.S., Schweitzer, L.: Changing demographics and the shifting nature of careers: implications for research and human resource development. Hum. Resour. Dev. Rev. **13**(2), 181–206 (2014)
24. Loogma, K.: How workers cope with changes in working life: acceptance strategies. In: Stenström, ML., Tynjälä, P. (eds.) Towards Integration of Work and Learning: Strategies for Connectivity and Transformation, pp. 173–185. Springer, Dordrecht (2009). https://doi.org/10.1007/978-1-4020-8962-6_10
25. Sato, T., Jackson, A., Yamani, Y.: Number of interrupting events influences response time in multitasking, but not trust in automation. Int. J. Aerosp. Psychol. **34**, 1–17 (2024)
26. Gibbs, M.B.: How is new technology changing job design? IZA World of Labor (2022)
27. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989)
28. Venkatesh, V., Davis, F.D.: A theoretical extension of the technology acceptance model: four longitudinal field studies. Manage. Sci. **46**(2), 186–204 (2000)
29. Venkatesh, V., Thong, J.Y.L., Xu, X.: Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q. **36**, 157–178 (2012)
30. Turban, E., Leidner, D., McLean, E., Wetherbe, J.: Information Technology for Management (with CD). Wiley, Hoboken (2008)
31. Fernández-Portillo, A., Almodóvar-González, M., Hernández-Mogollón, R.: Impact of ICT development on economic growth: a study of OECD European Union countries. Technol. Soc. **63**, 101420 (2020)
32. Meissner, A., Trübswetter, A., Conti-Kufner, A.S., Schmidtler, J.: Friend or foe? Understanding assembly workers' acceptance of human-robot collaboration. ACM Trans. Hum.-Robot Interact. (THRI) **10**(1), 1–30 (2020)
33. Hu, S., Laxman, K., Lee, K.: Exploring factors affecting academics' adoption of emerging mobile technologies—an extended UTAUT perspective. Educ. Inf. Technol. **25**, 4615–4635 (2020)
34. Sharma, R., Singh, G., Sharma, S.: Modelling internet banking adoption in Fiji: a developing country perspective. Int. J. Inf. Manage. **53**, 102116 (2020)

35. Alnemer, H.A.: Determinants of digital banking adoption in the Kingdom of Saudi Arabia: a technology acceptance model approach. Digit. Bus. **2**(2), 1000ban (2022)
36. Pang, C., Collin Wang, Z., McGrenere, J., Leung, R., Dai, J., Moffatt, K.: Technology adoption and learning preferences for older adults: evolving perceptions, ongoing challenges, and emerging design opportunities. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pp. 1–13 (2021)
37. Hecker, I., Spaulding, S., Kuehn, D.: Digital skills and older workers: supporting success in training and employment in a digital world. Urban Institute (2021)
38. Evers, M., Ryan, P.: As barristers embrace technology, it is a brave new world for their clerks. Aus. Bar Rev. **42**(3), 350–366 (2016)
39. Oude Mulders, J., Henkens, K.: Employers' adjustment to longer working lives. Innov. Aging **3**(1), igy040 (2019)
40. Landers, R.N., Marin, S.: Theory and technology in organizational psychology: a review of technology integration paradigms and their effects on the validity of theory. Annu. Rev. Organ. Psych. Organ. Behav. **8**(1), 235–258 (2021)
41. Eriksson-Zetterquist, U., Lindberg, K., Styhre, A.: When the good times are over: professionals encountering new technology. Hum. Relat. **62**(8), 1145–1170 (2009)
42. Barley, S.R.: Why the internet makes buying a car less loathsome: how technologies change role relations. Acad. Manage. Discov. **1**(1), 5–35 (2015)
43. McFarland, L.A., Ployhart, R.E.: Social media: a contextual framework to guide research and practice. J. Appl. Psychol. **100**(6), 1653 (2015)
44. Hughes, C., Robert, L., Frady, K., Arroyos, A.: A hiring paradigm shift through the use of technology in the workplace. In: Managing Technology and Middle- and Low-skilled Employees: The Changing Context of Managing People, pp. 49–59. Emerald Publishing Limited (2019)
45. Gong, B., Tobias, P., Young-Bristol, J.: Leveraging resources to improve supervisors' vision in the remote workplace. Manag. Res. Rev. **46**(6), 777–789 (2023)
46. Milić, D.C., Lukić, R., Delić, A.: Beyond the golden years: exploring continued workforce engagement in retirement. S Bus. A **81** (2024)
47. Zulfiqar, S., Younus, S., Khan, M.: Socio-Demographic factors influencing technology adoption in developing economies: a comparative study. J. Innov. Knowl. **7**(1), 123–135 (2022)

A Study on the Effects of Social Media Platforms on Graduate Students' Academic Achievement

Jatinkumar B. Kotadiya(✉) and Amit K. Patel

Computer Science Department, Atmiya University, Rajkot, India
jatin.kotadiya123@gmail.com

Abstract. The increasing ubiquity of social media platforms has changed the system of post-secondary education and has profoundly affected the lives of users, with almost everyone benefiting from it. It has also innovated the way student access data and connects with each other. The rise of online communities and the essence of the online environment have given rise to worldwide social networking platforms. The event includes numerous activities, from making virtual friends to chatting with them and exchanging images and videos. This research investigates the overuse of social media platforms and specifically the impact on the academic careers of students. The primary goal of this research is to target urban and rural graduate students in the Department of Computer Science at Harivandana College to determine how social media affects their academic performance and to examine both the strengths and weaknesses of social media use to reduce its use and improve one's academic career while providing suggestions for enhancing learning. An analytical research approach was used in this study, as well as a cross-sectional survey and questionnaire to collect data. The findings show that social media significantly affects a student's academic path. There are also physical and mental effects that can result. In addition, study hours may deteriorate. On the contrary, online earning through social media can increase one's income source. Notably, Instagram, WhatsApp, and YouTube emerge as the most used platforms for educational purposes among the surveyed students. This research sheds light on the multifaceted role of social media in the lives of college students and attempts to enhance academic activities. Social media and education are becoming increasingly entwined. It has also become an addiction that creates a gap between parents and students. Students prefer to share their problems, feelings, and frustrations with their elders on social media, as they find it easier. Such research enhances education and knowledge. The findings of my study will also be useful for educational policymakers and social networking site developers.

Keywords: college students · educational impact · social media · Academic Performance · learning experiences · E-learning

1 Introduction

With the revolutionary advancement of information technology, social media through smartphones has reached almost all people, including students. It has created ample opportunities and scope for students to interact, get information, enhance learning, and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 363–378, 2025.
https://doi.org/10.1007/978-3-031-86305-9_25

engages in problem-solving. As a highly interactive platform, social media helps students post their problems, get solutions, and analyze the different answers and opinions received, thus enhancing their imagination and creative skills [1]. They can easily and always communicate with their friends, classmates, and teachers through social media. Time and money can be saved if students use social media carefully. However, the use of accountable parenting is crucial to minimize potential negative effects and maximize gains in learning. Researchers highlight social media's function in bridging formal and informal knowledge approaches and emphasize the importance of avoiding addiction to social media use to achieve optimal results [2].

As the world has widely adopted social media in education, social media platforms such as WhatsApp, Facebook, Instagram, Snapchat, LinkedIn, YouTube, blogs, Skype, Twitter, Pinterest, Google Docs, and Wiki among youth and students have shaped information sharing and socialization in modern social interactions. While excessive use of this can lead to addiction and negative effects, judicious use of social media in conjunction with education can enhance students' academic careers as well as assist in social media exploration and social engagement [3].

Social media supports learning, enhances communication, and promotes academic success, ethics, and creativity. Researchers advocate for teachers to foster a supportive environment, plan activities comprehensively, and encourage communication. Using new approaches and activities to learn online makes the educational process easier [4-6].

A study has highlighted the impact of smartphones on academic performance, noting the benefits but also the increased risks of overindulgence in social media. Stress and depression may result from this, as well as hinder learning. Social media also interferes with the timely submission of assignments, access to lecture notes, and many academic activities. Hence, this research examines how social media affects student learning, acting as a mediator with technology self-efficacy and aiming to enhance academic outcomes through effective technology integration [7].

Education is a systematic process that promotes knowledge, skills, and growth. Globalization has given rise to education being integrated with information and computer technology (ICT). This new era has come to be known as the "Digital Age." This digital integration benefits students, parents, and professionals in various fields. Also, it serves as an invaluable asset [8]. Digital education has improved students' learning as well as their social engagement on various platforms. Rich and developed nations make the best use of ICT.

Early social media served as a platform for personal connections between family and friends. However, nowadays, colleagues are also gathering to use it for advertising, videos, sales, and recruitment. Despite its benefits, the influence of social media extends to activism, often in favor of its positive impact [9].

Many studies have proven that social media should be used to advance education, as it has positive effects on attitudes toward learning, academic success, and online knowledge sharing, as well as promoting creativity [10]. On the other hand, experts pointed out that the use of social media by students is often not officially integrated into educational programs. However, students also use it to collaborate on academic work such as project work, group assignments, case studies, etc. [11].

2 Related Work

The study concludes that self-regulation moderates the relationship between social media use and academic outcomes, suggesting further validity through structural equation modeling. This approach offers a model for analyzing how social media affects academic performance by fusing communication theory and TAM, with self-regulation as a mediator. Findings show favorable outcomes of perceived usefulness, ease of use, motivation, and ease of communication on social media use, which influence academic performance [12].

The study surveyed 505 students from a Chinese higher vocational college to determine their motivation to use social media for entrepreneurship education. These findings show a positive perception of the effectiveness of social media in collaborative learning but a low view of its utility in identifying entrepreneurial opportunities. Gender and family were shown to be factors influencing entrepreneurial attitudes, suggesting the need for additional study to optimize Social media's function in identifying opportunities and exploring demographic influences in more detail [13].

A survey conducted among communication officers in 55 French inland regional nature parks revealed problems hindering the use of digital technologies for the parks' educational purposes. Problems include a lack of confidence in providing online content and an underutilization of the potential of social media to promote environmentally responsible behavior. Recommendations emerged indicating the need for digital strategy implementation, resource allocation, collaboration promotion, and the establishment of clear educational goals. These findings were shared throughout the park network, prompting further research and interviews to address the identified issues [14].

This study conducted a focus group with 10 UAE university students to investigate the influence of social media on health and beauty brand awareness and their purchasing behavior. Questions were developed in collaboration with the brand, covering brand awareness, social media usage, and purchase patterns. Data analysis revealed insights into participants' brand awareness, preferences for health and beauty products, the influence of social media and influencers on purchase decisions, and the influence of favored social networking sites. The results highlight the significant influence of social media and influencer marketing on millennial purchasing power, urging brands to develop targeted strategies to reach this influential consumer segment [15].

This paper investigates how social media affects societies and businesses around the globe. Through a literature review, it examines benefits like audience engagement with privacy concerns and drawbacks such as cyberbullying. While social media fosters informal connections and business opportunities, it is important to address ethical standards and regulations to effectively navigate its risks [16].

This study used the UTAUT model to examine undergraduate students' perceptions of using social media for education during the COVID-19 pandemic at the University of Delhi, India. Data were collected from 279 students through a web survey, showing that performance expectancy, effort expectancy, facilitating conditions, and COVID-19-induced social isolation significantly influenced behavioral intention. The study also identified age, gender, and area of residence as significant moderating variables. Recommendations called for expanding the research to other universities, including education level and additional demographic variables for better generalizability [17].

Nowadays, students are becoming very sensitive due to the excessive use of social media and smartphones. A negative correlation is also being established between social media and students, as it harms a student's academic performance, career, and social behavior. We can say that social media has pervaded the student body. Getting addicted to social media and smartphones leads students to addiction. As we can see, if students are seeking informational content from social websites, at the same time, they will be in contact with other people through social media [18].

Students who find traditional teaching methods boring lose interest in studies, whereas, due to the active participation of students in social media, they increase in purely cognitive learning and show active participation in the learning process [19].

This study by [10] showed that people who spend more time inertly on social media suffer the most from loneliness, hopelessness, perceived inferiority, and depression.

When a person is exhausted from social media, there are more chances of increasing stress, anxiety, and even depression in his life. In other words, anxiety and depression are bad emotions that grow when a person is under stress [11].

3 Research Methodology

An analytical approach was used in the research conducted. An online questionnaire (Google Form) was utilized in a cross-sectional survey to gather data. Subsequently, a critical evaluation was performed on the gathered numbers and facts. The survey members were students, and the purpose was to determine their perspective on the use of social media in learning. This methodology enables a comprehensive examination of the data to reveal patterns, relationships, and insights pertinent to the research objectives.

For this study, there were graduate students from urban and rural areas in the Computer Science department at Harivandana College. The sample of 160 students was selected using purposive sampling, which involves selecting a specific group or class of students to represent the entire group based on characteristics. Specifically, students from the college's undergraduate program, aged 17 to 22, were included in the sample, reflecting the study's interest group. The survey revealed that all institute students have internet access from their residences. While the traditional classroom is the primary learning method, the school also utilizes educational platforms like Google Classroom for additional support. Additionally, students are familiar with various online social networks outside their academic environment and actively engage in exchanging information.

This study looks at how students use social media to fulfill particular needs. Some theory was used to understand it. This theory posits that individuals are active participants in their social media use, selecting and interpreting content based on their needs and preferences. It asserts that people use social media to satisfy their specific desires, suggesting that social media also competes with other sources of information to meet users' need [20].

In this study, two steps were first taken to construct the questionnaire and its validation process. Initially, an expert evaluation was done by three faculty members in the respective field to establish the validity of the questionnaire. Internal consistency was then assessed by administering this questionnaire to a randomly selected group of 10 college students, resulting in a good Cronbach alpha score of 0.85.

4 Findings

The goal of this research was to examine the correlation between students' use of social media and their academic achievement, as well as to investigate specific variables related to the findings and gratifications derived from these platforms. These variables were selected to clarify social media's effects on academic outcomes as well as to explore potential mechanisms in online and offline learning patterns arising from these findings.

4.1 How do Social Network Variables Such as the Number of Connections, Frequency of Interactions, and Type of Content Shared Relate to Student Academic Performance?

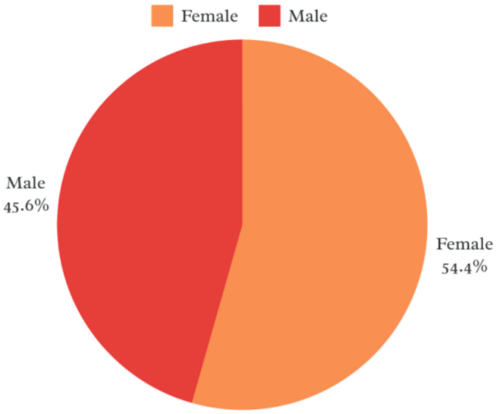


Fig. 1. Use of Social media by students' gender

As shown in Fig. 1, the more time adults currently spend communicating, the more time young people prefer to be active on social media. A sample of 160 students was taken, with 54.4% girls and 45.6% boys. Girls responded more than boys in this survey.

As shown in Fig. 2, it is somewhat difficult to answer who does not use social media these days. Everyone, from small children to old people, is covered on social media. At such a time, it is something to think about how essential social media is for children. In this survey, the average age of the students was 18 to 21 years. 10% of students are 18 years old, 20% are 19 years old, 40% are 20 years old, and 45% are 21 years old.

As shown in Fig. 3, young people are constantly wasting time on social media. Not only this, but social media is also damaging their mental health. Which is becoming a big threat to the youth as well as the future of India? Then, in this survey, 46.9% of students in FY, 21.9% of students have S.Y. Also, 31.3% of students in T.Y.

As shown in Fig. 4, social media has changed the world. While everyone from rural areas to urban areas uses social media on the same platform, in this study, 48.1% of students were from rural areas and 51.9% of students were from urban areas.

As shown in Fig. 5, students use more than one social media account to accomplish educational goals as well as share information. Also, students can learn a lot of new

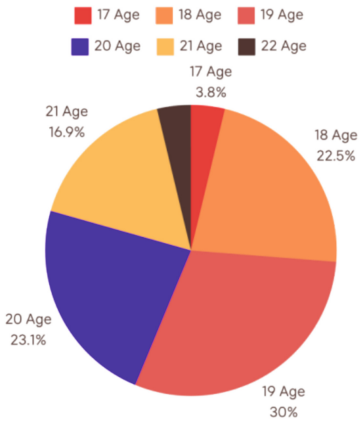


Fig. 2. Use of Social media by students' Age

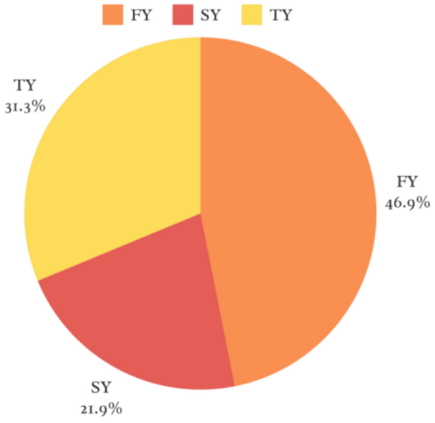


Fig. 3. Use of Social media by students' course

things through social media. You can easily keep yourself and others updated with news, world happenings, political information sharing, messaging, etc. through social media. The report found that the highest number of students, 25%, have only 1 social media account, while 23.1% have 2 social media accounts. The number of students having three social media accounts was 23.1%, and the number of students having more than three social media accounts was 28.8%.

As shown in Fig. 6, excessive use of social media floods our brains with negativity, sometimes leading us to depression. Its excessive use damages physical and mental health. Have students tried to find out if the use of social media affects their studies? It was then found that the highest 49.4% of students believe that social media has a serious impact on education, while 18.8% disagree and 21.3% believe that it can. From this, it can be seen that students are aware of the impact of social media on education.

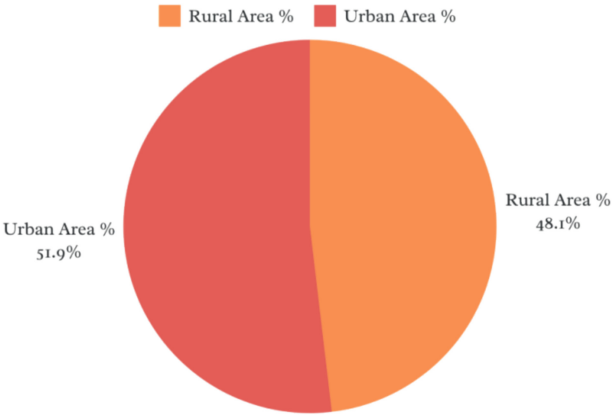


Fig. 4. Use of Social media by students' geographic areas

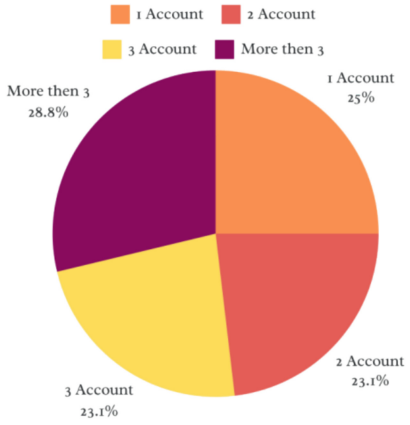


Fig. 5. How Many Social Media Account used by Student's

As shown in Figure 7, in the current digital era, where social media has brought the world closer together, the overuse of various devices to provide information sharing has health implications. Many people sit in their rooms alone for hours in front of a computer, mobile device, or TV. They do not get the physical exercise they need. So a sedentary life can lead to heart problems, diabetes, and other major diseases. For that, the students tried to know whether social media could affect their health. It was found that the highest 49.4% of students believed that excessive use of mobile phones or other devices can have serious effects on health, while 16.9% disagreed, 18.1% of students believed that it may be, and 15.6% of students believed that this matters. But this is not an idea; it can be known from this that students are aware of the effect of social media on health.

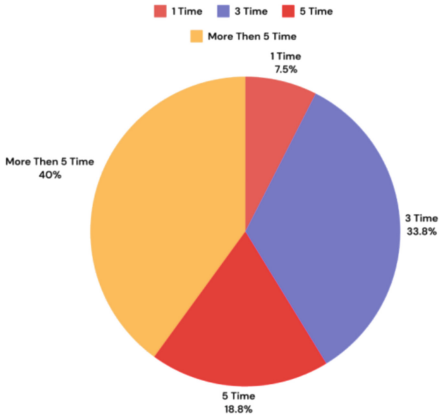


Fig. 6. Social media affects Student's education

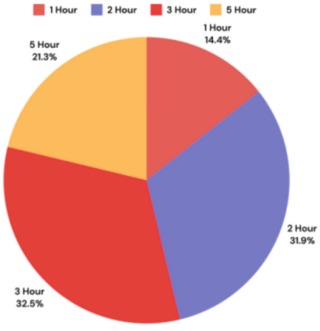


Fig. 7. Social media affects student's health.

4.2 What are the Primary Uses of Social Networks Among College Students, Including Communication, Information Sharing, Entertainment, and Socializing?

As shown in Fig. 8, what did students report as their primary use of social media? In today's internet age, when social media is used for different purposes, it was found that 41.3% of students use social media for information sharing, 35% for entertainment, 15.6% for communication, and 8.1% for socializing. Students mostly use social media for self-generated content such as text, posts, comments, digital photos, and videos.

As shown in Fig. 9, social media used by students for educational purposes include WhatsApp, Facebook, Instagram, LinkedIn, YouTube, blogs, Skype, Twitter, Pinterest, and Telegram. Blogs were the least used of all websites. Students found it easier to use Instagram (84.4%) and WhatsApp (79.4%), while 70.6% used YouTube to learn new things by watching videos.

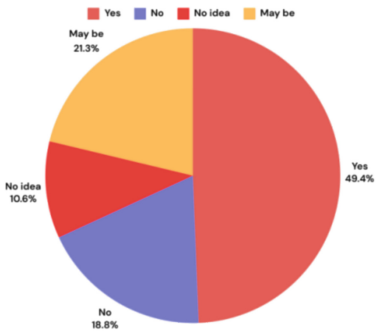


Fig. 8. Primary uses of social media by students

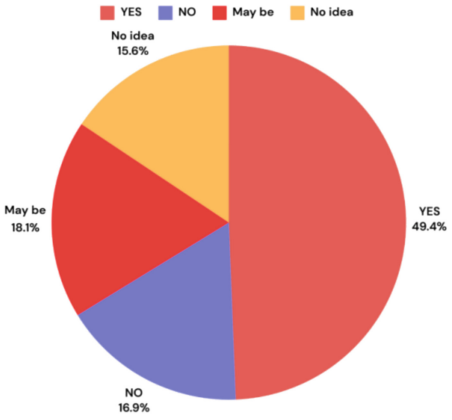


Fig. 9. Social Media is widely used by students.

4.3 How do Students Utilize Social Networks on Their Mobile Phones, and which Specific Features or Functionalities are Most Commonly Accessed?

As shown in Fig. 10, while there are different devices available to use on many social media platforms, 78.8% of students reported that they use social media most especially on their smartphones. While 9.4% use laptops or desktops as well as systems, the increasing numbers of mobile students are also using social media platforms for different visions, which provide location, photos, videos, and many more. Due to the widespread presence of smartphones, social media applications on mobile are becoming more widespread, while global mobile social media usage shows no signs of slowing down.

4.4 What is the Average Amount of Time Spent by Students on Social Networking Sites Per Day, and How does This Usage Pattern Vary Across Different Demographic Groups?

About 40% of the students said they use social media more than five times a day, as illustrated in Fig. 11. This increased my flexibility to stay in touch. The popularity

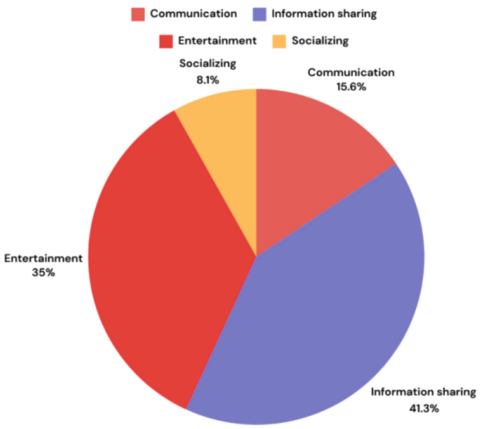


Fig. 10. Devices used by students to access social media.

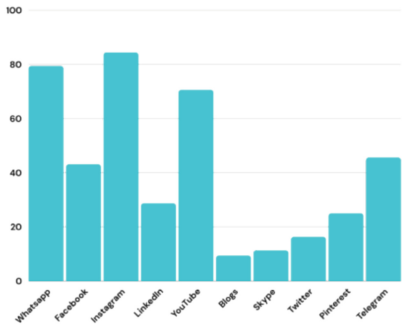


Fig. 11. Social media is used by students throughout the day.

of social media has increased due to the smart features available on social networks, like image sharing, video sharing, and status updates. This was not directly related to education. While 33.8% of students use it three times, 18.8% use it five times, 40% use it more then, five times and the lowest 7.5% of students use social networking sites only once during the day. From these findings, it can be said that this is a good thing.

As shown in Fig. 12, the modern age is one of technology. Everyone wants to be social, whether they are social or not. In other words, a person feels incomplete without social media. Today, social media is now a significant means of connecting with people around the world and has given a new direction to world communication. When students were asked the number of hours a day they spend on social media, the answers were shocking: 14.4% spent one hour, 31.9% spent 2 h per day, 32.5% spent 3 h per day, and the highest 21.3% spent 5 h per day. A consumption level of more than three hours is the highest.

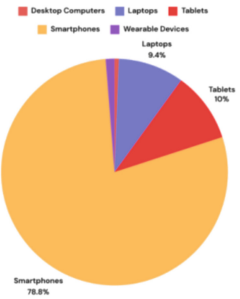


Fig. 12. Students spend time on social media during the day.

4.5 What are the Diverse Uses of Social Networking sites Among Students, Including Networking, Educational Purposes, Leisure Activities, and Professional Development?

Fig. 13. Special features used by students on social media

As shown in Fig. 13, what specific features do students use on social media? Students were then able to show the various purposes for which they used social media. In Fig. 13, I see 71.3% for messaging and chat, 67.5% for file sharing, 26.9% for announcements and updates, 33.1% for personalized learning recommendations, 18.1% for live streaming and webinars, 23.1% for discussion forums, and 25% for blogs and journals. The least used features were polls and surveys at 18.1%.

As shown in Fig. 14, the era of social media is now underway. Every second, a person is seen creating a reel on social media. People from young to old are now using social media, and most people are active on Instagram. The report found that the highest use of social media by students, at 70%, is for downloading and uploading music and videos, while at least 15.6% is for Collaborative Document Editing.

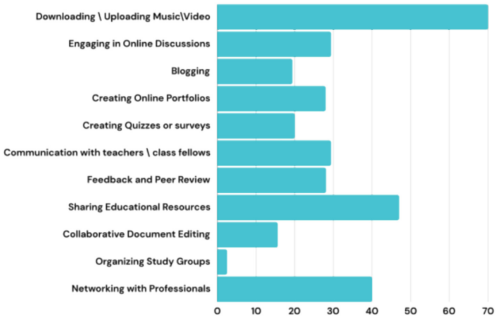


Fig. 14. Special features used by students on social media

5 Discussion

This study used a cross-sectional survey, questionnaires, and an analytical research approach to collect data. The main objective of this study was on urban and rural graduate students of the computer science department at Harivandana College, where 160 students were collected through purposive sample records. The purpose of the study was to determine how social media factors and students' academic achievement related to one another. It included male and female students in their first, second, and third years of college, aged between 18 and 21. The study shows that students use many social media accounts, with 20% using more than 3 accounts, 50% using them more than 5 times a day, and 33% spending more than 5 h a day. Excessive technology use hurts both education and health. The study also found that 90% of students believe social media impacts education, while 40% believe it seriously impacts health. Nonetheless, they are aware of it.

The findings of this study show that various aspects of social network usage among students affect academic achievement. Notably, students mentioned using different social media platforms, for instructional purposes, such as Facebook, YouTube, WhatsApp, and Instagram. While blogs and Twitter were less frequently used, Facebook, Instagram and WhatsApp were used because of their simplicity of use. Approximately 78.8% of students said they used social networking sites on their mobile devices, demonstrating that applications on these platforms are easy to access and use. The most often used smart features on mobile devices were location tagging, sharing of images and videos, and status updates; however, these features had nothing to do with education. This implies that students' use of social media is incorporated into their academic lives, which may have a greater impact on their learning.

The findings of this study are similar to those of [21], which show that social media is the most important tool for academic engagement among students, supportive peers, and teachers to advance their academic careers.

However, [22] a study found that excessive use of social media could not only impact students' focus on studies but also negatively affect their academic performance. Another study also noted that excessive use of social media by students in college may be harmful to academic achievement [23]. Additionally, it was observed that, due to excessive usage of social media, students tend to avoid their academic obligations and spend their time

there. On the other hand, a different study suggested that students' excessive use of social media increases their inability to concentrate and recall facts when they regularly utilize social media alongside academic tasks [24].

For what purpose do students use social media? Also, the time they spend is worth it. 32.5% of students spend more than three hours a day are spent. This indicates that a noteworthy part of their day is wasted on social networks, which poses a threat to their academic career. When students spend their time on social media, the possibility of negative consequences for their academic performance increases. What remains to be seen is how students use social media. Their effects may also change according to personal preferences. These results are similar to [22], which demonstrated how using social media excessively cuts down on study time and harms students' academic careers. The study highlighted the potential correlation between students' social media usage and their academic achievement.

Students use social media for various purposes, with its primary uses being communication (15.6%), information sharing (41.3%), entertainment (35%), and socializing (8.1%). These results suggest that while a few students utilize social media. solely for sharing pictures and videos and just for fun, they also participate in instructive exercises. The collaborative and interactive social media's nature can enhance students' academic professions by offering significant educational opportunities.

Remarkably, all students admitted to utilizing social media. Like Google Drive, WhatsApp, and YouTube the most while writing classroom assignments. This social media helps students get information and guidance related to their subject. [25] Also noted is that social media has a multifaceted role in education, with the most common use for homework purposes.

A noteworthy 83.8% of pupils said they were okay with social media. Zoom and Google Meet as effective e-learning tools, improving learning and sparing students' time. The report offers a number of social media applications in the educational context, such as forming distinct groups for every topic, Encouraging teachers to communicate, and sharing information with classmates [25]. Also added in their report that social media can perform well, be used effectively for education, and be career-oriented if used properly [23]. It has been found that usage of social media has increased is negatively associated with academic achievement. While these possible difficulties suggest how social media can be made effective in e-learning, understanding them is also a challenge.

As you know, social media has special features like discussion forums, messaging and chat, file sharing, announcements and updates, polls and surveys, live streaming and webinars, blogs and journals, and personalized learning recommendations, which are advancing day by day. The study also found that the social media most used by students are messaging and chat (71.3%) and file sharing (67.5%), while 70% of students use it for downloading or uploading music or videos, which may prove dangerous for their academic work.

6 Conclusion

In conclusion, the primary goal of this research is to determine the perspective from which graduate students in urban and rural areas, in the computer science department of a college in Harivandana, perceive and utilize social media for educational purposes.

Social media plays a very important role in education nowadays. We found that most students are using social media on their smartphones, with Instagram and WhatsApp being the most commonly used apps. The majority of students use social media for learning for 1 to 3 h daily.

In this study, we focused on students to acknowledge their perspective on the use of social media in education. Various factors have been surveyed on the positive and negative aspects of academic performance. On the positive side, teaching has led to self-paced learning. Then we can learn new things and explore career opportunities anytime and anywhere. In the same way, everything comes with negative aspects. Using social media also leads to many serious health problems that we cannot deny. To address these issues, counseling programs can be conducted to educate students about the potential mental health risks of social media for their health. As our saying goes, there are two sides to a coin: one negative and one positive. However, according to the respondents in this study, we found that the positive attitude outweighs the negative. So we can say that the use of social media in education has greatly benefited students. However, Social media use needs to be controlled because overuse can interfere with academic work.

7 Future Work

This study has several limitations that must be acknowledged. Most studies on the possible outcomes of social media use on students' psychological well-being are either in terms of numbers or words. The same research design with students of the same academic standing and the same age groups may have varying effects and related findings in different countries. The potential impact of popular social media can be both positive and negative. Thus, different contexts and further research are needed to compare the results. Further research is likely to be carried out in the future to show how the ongoing use of social media by students in school or university can affect the fluctuation of any affective components.

A sample size of 160 students has been taken in this study, which is a constraint. Due to the possibility that a sample size this small may not be entirely representative of all college or university students, it becomes difficult to generalize results to a broader framework. Social media's impact would be better explained by a larger and more varied sample of scholastic achievement. Second, this study only looked at one college. Students at another college may have their own unique personality and culture, and the findings of this institution may not apply to another college or university. As the student population increases, results may vary, and not all the data collected may be completely accurate in relation to the actual impact of social media on student academic achievement. If future research were conducted in different colleges or areas for educational purposes, Comparative studies looking at differences in social media use would greatly improve the outcomes. Future research could also look at how social media usage impacts academic achievement in particular subject areas or fields. Whether social media is usefulness or destructiveness varies among topics, which is why these findings need further research.

8 Recommendations

Social media helps you maintain long-term connectivity with your relatives, no matter where you are in the world. Thus, no matter where the teacher is, they can connect with the students. Teachers can also guide students by sharing materials, pictures, messages, and videos to help them achieve their academic objectives. With the help of social media, students can know the details of any educational events from home and participate in developing their skills. It also connects the child's parents with school activities, so they can easily check their child's progress by attending school meetings.

Considering the conclusions and debates from this study, multiple suggestions are also made for upcoming studies to improve social media use in the classroom in today's technological era. Successful implementation of social media in education requires developing and assessing privacy tactics to bridge the digital gap with equal availability to every student. In conclusion, efforts are needed to examine and understand the long-term effects of colleges and universities integrating social media into education and how it influences the academic achievement, engagement, and success of students.

References

1. Devi, K.A., Karim, Md. R., Sodhi, K.K., Bansal, R.: A study on student's perspective on use of social media in education. In: 2021 International Conference on Computational Performance Evaluation (ComPE), pp. 682–685 (2021). https://doi.org/10.1109/ComPE53109.2021.9751858
2. Siddiqui, S., Singh, T.: Social media its impact with positive and negative aspects. Int. J. Comput. Appl. Technol. Res. **5**, 71–75 (2016)
3. Shanaa, Z.A., Naser, K.M., Abualrish, M., Zaitoun, E.A., Yousef, J.: Effects of social networking sites on the academic performance of graduate students. In: 2023 Tenth International Conference on Social Networks Analysis, Management and Security (SNAMS), pp. 1–6 (2023). https://doi.org/10.1109/SNAMS60348.2023.10375420
4. Khovrak, I., Chernenko, S., Khovrak, I.: Exploring the role of social media platforms in fostering collaborative learning in distance education. In: 2023 IEEE 5th International Conference on Modern Electrical and Energy System (MEES), pp. 1–5 (2023). https://doi.org/10.1109/MEES61502.2023.10402504
5. Ansari, J.A.N., Khan, N.A.: Exploring the role of social media in collaborative learning the new domain of learning. Smart Learn. Environ. **7**, 9 (2020)
6. Liu, S., Zaigham, G.H.K., Rashid, R.M., Bilal, A.: Social media-based collaborative learning effects on student performance/learner performance with moderating role of academic self-efficacy. Front. Psychol. 13 (2022)
7. Nand, S., Shaikh Ali, S.H.: Impact of excessive use of social media on students learning performance: Gratifications theory perspective. In: 2022 International Visualization, Informatics and Technology Conference (IVIT), pp. 35–42 (2022). https://doi.org/10.1109/IVIT55443.2022.10033346
8. Ishaq, K., Zaki, F., Abid, A., Abid, K., Ijaz, M.: Impact of social media on students' academic performance & generation gap: a study of Public Sector University in Punjab. In: 2019 International Conference on Innovative Computing (ICIC), pp. 1–9 (2019). https://doi.org/10.1109/ICIC48496.2019.8966688
9. Sharma, A., Kaushal, S., Fulzele, R., Kumari, T.: Role of social media in engaging students: perspectives from higher education. In: 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 1683–1686 (2022)

10. Aalbers, G., McNally, R., Heeren, A., de Wit, S., Fried, E.: Social Media and Depression Symptoms: A Network Perspective. In press, J. Exp. Psychol. Gen (2018)
11. Iwamoto, D., Chun, H.: The emotional impact of social media in higher education. Int. J. High. Educ. **9**, 239 (2020)
12. Rfeqallah, M., Kasim, R., Al-Sharafi, M.A.: Conceptualizing a model for using social media as a learning tool and its effect on academic performance: the moderating effect of self-regulation. In: 2021 International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IoE), pp. 1–6 (2021). https://doi.org/10.1109/ITSS-IoE53029.2021.9615292.
13. Chen, L., Ifenthaler, D.: Mobile learning and social media applied in entrepreneurship education courses in a higher vocational college. In: 2021 IEEE International Conference on Engineering, Technology & Education (TALE), pp. 1023–1027 (2021). https://doi.org/10.1109/TALE52509.2021.9678928
14. Di Chiacchio, L., Garcia-Perez, A., Jarvis, D.: Setting a strategy for the use of social media as an educational tool to promote eco-responsible behaviours: the case study of nature Parks in France. In: 2023 IEEE International Conference on Technology and Entrepreneurship (ICTE), pp. 122–127 (2023). https://doi.org/10.1109/ICTE58739.2023.10488521.
15. Sayyed, B.J.W., Gupta, R.: Social media impact: generation z and millenial on the cathedra of social media. In: 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), pp. 595–600 (2020). https://doi.org/10.1109/ICRITO48877.2020.9197995.
16. Minocha, L., Jain, P., Singh, A., Pandey, P.: Social media's impact on business and society: a study. In: 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 2078–2081 (2022)
17. Nain, H.: Understanding students' behavioural intentions to use social media learning in education: a utaut model approach. In: 2021 International Conference on Computational Performance Evaluation (ComPE), pp. 256–261 (2021). https://doi.org/10.1109/ComPE53109.2021.9752450
18. Giunchiglia, F., Zeni, M., Gobbi, E., Bignotti, E., Bison, I.: Mobile social media usage and academic performance. Comput. Hum. Behav. 82 (2018)
19. Rueda, L., Benitez, J., Braojos, J.: From traditional education technologies to student satisfaction in Management education: a theory of the role of social media applications. Inf. Manage. **54**, 1059–1071 (2017)
20. Omar, F.I., Sallehuddin, I., Hafizah, N., Hassan, N.A., Gen, Y.: A STUDY ON SOCIAL MEDIA USE AND OUTCOMES (Gen Y: KajianPenggunaan Media SosialdanHasil). E-J. Manag. Muamalah 6 (2016)
21. Chen, M., Xiao, X.: The effect of social media on the development of students' affective variables. Front. Psychol. 13 (2022)
22. Azizi, S.M., Soroush, A., Khatony, A.: The relationship between social networking addiction and academic performance in Iranian students of medical sciences: a cross-sectional study. BMC Psychol. **7**, 28 (2019)
23. Tafesse, W.: Social networking sites use and college students' academic performance: testing for an inverted U-shaped relationship using automated mobile app usage data. Int. J. Educ. Technol. High. Educ. **19**, 16 (2022)
24. Shahid, D.R., Ijaz, M.U., Murtaza, M.S., and G. Impact of social media on academic performance of students: a survey conducted in urban and rural areas of Sahiwal. J. Posit. Sch. Psychol. **7**, 651–663 (2023)
25. Sharma, S., Behl, R.: Analysing the impact of social media on students' academic performance: a comparative study of extraversion and introversion personality. Psychol. Stud. **67**, 1–11 (2022)

Enhancement of Forest Fire Assessment by KDCPMNN Approach in Sikkim, India Using Remote Sensing Images

Kapila Sharma(✉) and Gopal Thapa

Sikkim Manipal Institute of Technology, Sikkim Manipal University, Gangtok, India
kapila.s@smit.smu.edu.in

Abstract. Forest fires pose a significant threat to ecosystems, especially in regions like Sikkim, where climate conditions increase the frequency of forest fire. Accurate prediction and assessment of forest fires is essential to mitigate damage to forest land and wildlife. However, existing research has not adequately combined meteorological data with land imagery for forest fire prediction in this region. This study proposes a novel Kernelized Deep Convolutional Parametric Modswish Neural Network (KDCPMNN) model for forest fire prediction using satellite images and meteorological data. Initially, satellite images are segmented using the K-Means (KM) algorithm, followed by pre-processing with Ani-sotropic Exponential Diffusion Filter (AEDF) and Bayesian Contrast Limited Adaptive Histogram Equalization (B-CLAHE). The Dark Channel Prior (DCP) and Transmission Map (TM) are then used to calculate intensity values and depth information, respectively, from which features are extracted. Simultaneously, vegetation details and attributes from meteorological data are also extracted. These combined features are input into the KDCPMNN to predict the likelihood of a forest fire. In the event of a fire, the burn area is assessed using the Composite Burned Index (CBI) within a Secant Linear Membership-based Fuzzy Inference System (SLM-FIS), while the fire's potential is evaluated through confidence score. Experimental results show that the proposed KDCPMNN model achieves 98.2% accuracy, surpassing existing methods.

Keywords: Secant Linear Membership based Fuzzy Inference System (SLM-FIS) · Kernelized Deep Convolutional Parametric Modswish Neural Network (KDCPMNN) · Anisotropic Exponential Diffusion Filter (AEDF) · Bayesian Contrast Limited Adaptive Histogram Equalization (B-CLAHE) · Dark Channel Prior (DCP) · Transmission Map (TM) · Composite Burned Index (CBI)

1 Introduction

Forest fires are increasingly recognized as one of the most frequent natural disasters worldwide [12], causing significant damage to ecosystems, wildlife, and human settlements. Major fire events, such as those in southeastern Australia in 2019–2020, resulted in considerable environmental and economic loss [7, 11]. In India, the state of Sikkim,

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 379–394, 2025.
https://doi.org/10.1007/978-3-031-86305-9_26

located in the high lightning-prone northeastern region, has also witnessed fire incidents triggered by natural phenomena like lightning. Given this vulnerability, the early prediction of forest fires is crucial for minimizing damage [3, 14]. Satellite imagery has proven to be highly effective for forest fire prediction, particularly in active fire states. One key factor in predicting forest fires is the extraction of biomass values from satellite images [5]. In addition, the Normalized Difference Vegetation Index (NDVI) provides essential information for assessing fire severity [4]. These parameters, derived from satellite data, are input into prediction models. Recently, machine learning (ML) methods have gained popularity for Forest Fire Detection (FFD) [8], with deep learning (DL) techniques, which efficiently capture spatial correlations in image data, becoming the leading approach [17]. Among these, Artificial Neural Networks (ANN), Generalized Boosting Machines (GBM), Random Forests (RFs), and Maximum Entropy (MaxEnt) are widely used for detection [6]. However, existing research still faces several challenges. No current studies have integrated meteorological data with land cover imagery for forest fire detection in Sikkim. Moreover, issues such as misprediction due to the similar visual characteristics of smoke and haze [13], ineffective management in assessing burn area and fire risk [1], and poor classification performance caused by image blurriness [23] persist. Accurate prediction and assessment of forest fire risks remain a critical challenge, especially in areas like Sikkim, India, which are highly vulnerable due to their unique topography and climatic conditions. To address these challenges, this study proposes a Kernelized Deep Convolutional Parametric Modswish Neural Network (KDCPMNN) designed specifically for predicting forest fire risks in the Sikkim region. The forest meteorological data, crucial for performance analysis, is sourced from Kaggle's dataset repository [Dataset: https://www.kaggle.com/datasets/elikplim/forest-fires-data-set], while Landsat OLI images are collected from Google Earth. By integrating these datasets, KDCPMNN model can capture critical vegetation and climate patterns, enhancing the accuracy of fire risk assessments. This study highlights the potential of combining remote sensing data with meteorological information to develop more robust forest fire prediction systems.

1.1 Objectives

- To include the meteorological data with the land cover image for forest fire prediction in Sikkim, India.
- To extract depth information and intensity values by TM and DCP.
- To assess the burning area and potential risk level of fire accidents utilizing the SLM-FIS approach.
- To enhance image quality by AEDF and B-CLAHE approaches.

2 Related Work

The Unmanned Aerial Vehicle (UAV) FFD system was suggested by [23], by integration of Artificial Intelligence (AI) approaches. To detect FFD, the research utilized different ML models. Regarding accuracy, the simulation outcome illustrated the presented research's efficacy. The UAV-based system focuses primarily on aerial data for fire detection but does not incorporate other critical data sources, such as satellite imagery and

meteorological data. An enhanced FFD model utilizing a DL approach was introduced by [13], the Detectron2 model was utilized by the model. The model's performance was analogized with the other models. Higher accuracy was achieved by the presented method. The research gap lies in the lack of multi-source data integration, region-specific adaptability, and comprehensive fire risk prediction in existing forest fire detection models. An ensemble learning model was introduced by [1], to detect forest fires in distinct scenarios. Two individual learners, namely Yolov5 and EfficientDet, were utilized in this research for the detection process. Experimental analysis was done for the presented research, and it provided superior performance than the conventional model.The study focuses on detection rather than integrating predictive capabilities using diverse data sources, such as meteorological and satellite imagery, and its lack of adaptability to region-specific fire risk scenarios.

A fire and smoke detection system were proposed by [20] employing an AI approach. As per experimental analysis, the model achieved better performance than the prevailing model. Nevertheless, when employing the conventional CNN model, the False Positive Rate (FPR) was high. However, high False Positive Rate (FPR) associated with the conventional CNN model, indicates a need for improved algorithms or methodologies that enhance detection accuracy while reducing false alarms, particularly in diverse environmental conditions. An approach for fire detection was presented by [22], on combining color-motion-shape features with machine technology. As per the experimental outcome, the method achieved a better true-positive rate and false-negative rate than the conventional model. The images obtained from the physical device were of poor quality, thus affecting the model's performance. Forest Fire susceptibility as well as management strategies was proposed by [9] in the Western Himalayas by integrating ensemble ML and explainable AI. By employing a stacking framework centered on the ensemble models with the DNNs as a meta-model, the performance and reliability of this model were enhanced. For improving this model's interpretability, the XAI was utilized. The analysis outcomes stated that for the forest fire susceptibility prediction, this system attained significant F1 values. But the lack of interpretability was a major problem in this model. Automatic processing model to identify Forest fire risk areas along high-voltage power lines by utilizing coarse-to-fine LiDAR data was developed by [10]. A coarse-to-fine technique was developed for the preliminary analysis of the airborne LiDAR data as well as to classify the fire risk areas. According to the evaluation outcomes, this model was expressively used for the high-fire risk areas selection. But this model generated a risk map only regarding the KDA. Spatiotemporal pattern analysis model in the burned areas centered on the GIS for fire risk monitoring was presented by [2] the burned areas' spatiotemporal patterns and the fire-prone areas' density were efficiently analyzed. Hence, this model's robustness in analyzing the forest and land fire risk was stated by the experimental analysis. Yet, this model reported a very minimal forest fire susceptible area.

3 Proposed KDCPMNN-Based Forest Fire Assessment in Sikkim, India

The forest fire is assessed utilizing the KDCPMNN approach by considering vegetation indexes and meteorological data. Figure 1 displays the structure of the presented research.

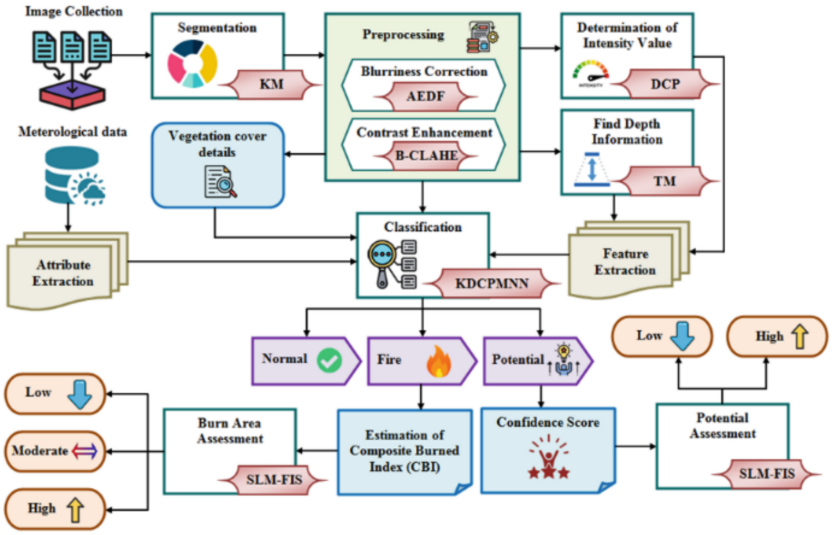


Fig. 1. Structure of the proposed methodology

3.1 Image Collection

The input images for the study are primarily sourced from the Landsat OLI satellite image gallery. These collected satellite images $\wp_c$ serve as the foundational dataset for the analysis.

3.2 Segmentation

To improve the quality of the image, forest land is segmented from $\wp_c$ by employing the KM algorithm. The KM algorithm is supported for huge quantities of data. Primarily, from $\wp_c$, it randomly chooses the centroids, which are notated as $\aleph_s$. After that, the difference $\Re_d$ between $\wp_c$ and $\aleph_s$ is computed utilizing Eq. (1),

$$\Re_d = \sqrt{\sum\left(\wp_c - \aleph_s\right)^2} \tag{1}$$

Subsequently, the pixels are grouped with similar differences. Here, $\hbar_{for}$ signifies the segmented forest area.

3.3 Preprocessing

To reduce errors and improve image quality, the following pre-processing steps are employed:

(a) **Blurriness correction**

By using AEDF, the blurriness of $\hbar_{for}$ is corrected. The conventional Anisotropic Diffusion Filter (ADF) effectively smoothens the image and preserves the edge details. Nevertheless, the parameters' adjustment is varied for each iteration, thus affecting the

filtering process. Thus, by using the exponential function, this research methodology considers the constant value for the image's gradient magnitude.

$$\Im_{blr}\left(\hbar_{for}\right) = \hbar_{for}(\beta) + \frac{\ell}{|\phi_\beta|}\sum \chi\left(\left|\nabla \hbar_{for}^{\beta,q}\right|\right)\nabla \hbar_{for}^{\beta,q} * exp(g_d) \tag{2}$$

$$\nabla \hbar_{for}^{\beta,q} = \hbar_{for}(\beta) - \hbar_{for}(q) \tag{3}$$

The blurriness corrected image is notated as $\Im_{blr}$, the pixel position is defined as β, the conductance function is symbolized as χ, the rate of diffusion is illustrated as ℓ, ϕ_β defines the spatial 4-pixel neighborhood of pixels β , the gradient operator is denoted as ∇, the exponential value of the magnitude gradient of the $\hbar_{for}$ is specified as $\exp(g_d)$ and the difference between the neighborhood pixels is signified as $\nabla \hbar_{for}^{\beta,q}$.

(b) **Contrast enhancement.**

In this methodology, the B-CLAHE algorithm is employed to enhance contrast level of $\Im_{blr}$. . While the conventional Contrast Limited Adaptive Histogram Equalization (CLAHE) method improves contrast by correlating the intensities of neighboring pixels, it tends to introduce noise and detail loss in areas with high brightness. To address this issue, the Bayesian probability function is applied to adjust the brightness intensity of the image. Initially, the regions of the image $\Im_{blr}$ are segmented and represented in Eq. (4).

$$\psi_{reg} = \{\psi_1, \psi_2, ..., \psi_n\} \tag{4}$$

where, the region of $\Im_{blr}$ is notated as ψ_{reg} and the n-number of regions is defined as. After that, the histogram of the contextual region is computed and summed up λ_{sum}, which is presented in Eq. (5). In addition, based on Bayesian probability, the clip limit is derived in Eqs. (6) and (7).

$$\lambda_{sum} = \sum \psi_{reg} \tag{5}$$

$$\varpi_{ag} = \frac{P_b(U_{dir}) * P_b(V_{dir})}{\Im_{blr}(o)} \tag{6}$$

$$P_b = \frac{P_b(V_{dir}|U_{dir}) \bullet P_b(U_{dir})}{P_b(V_{dir})} \tag{7}$$

The average pixel notated as ϖ_{ag}, $P_b(U_{dir})$ and $P_b(V_{dir})$ defines the clip limit probability of the pixels in U and V direction of the region, respectively, and the overall pixels in the image region are indicated as $\Im_{blr}(o)$. Later, the actual limit κ_{al} is described in Eq. (8),

$$\kappa_{al} = l * \varpi_{ag} \tag{8}$$

where, l denotes the clip limit and η_e represents the contrast-enhanced image.

3.4 Determination of Intensity Value

In this process, the intensity value is calculated for η_e to extract relevant information. The Dark Channel Prior (DCP) method is employed in this research to determine the intensity value.

$$\delta_{dark} = \min\left(\min \eta_e^c\right) \tag{9}$$

where, the DCP output is indicated as δ_{dark} and the local patches of the image are defined as η_e^c.

3.5 Estimation of Depth Information

The depth information η_e is estimated through the use of Transmission Map (TM) construction δ_{trans},

$$\delta_{trans} = A_t(\eta_e) \cdot \varepsilon(\eta_e) \tag{10}$$

where, the illumination value is represented as $A_t(\eta_e)$, and the light reflection rate is denoted as $\varepsilon(\eta_e)$.

3.6 Feature Extraction

Enhancement of performance was done for which, features are extracted from δ_{dark} and δ_{trans}. Color, Correlation, Kurtosis, Skewness, Speeded Up Robust Feature (SURF), etc., are the features. The extracted features are notated as γ_f.

3.7 Vegetation Cover Details

In this section, to improve the prediction performance, the Land Use Land Cover (LULC), NDVI, and Forest Density are obtained from η_e for considering vegetation properties. In addition, topography, Distance to Village, and Distance to Road are obtained from Google Earth of η_e, and the proximity maps are generated utilizing ArcGIS software. Here, μ_g defines the obtained vegetation detail.

3.8 Meteorological Data

Conversely, meteorological data is gathered to obtain details about climate and other properties of forest land. The collected data is notated as M_d.

3.9 Attribute Extraction

The attributes like a month, day, Fine Fuel Moisture Code (FFMC), Drought Moisture Code (DMC), Drought Code (DC), etc., are extracted from M_d. The extracted attributes are symbolized as α_e.

3.10 Classification

To predict forest fire efficiently, γ_f, μ_g, η_e and M_d are given as input to the KDCPMNN classifier. Conventional Deep Convolutional Neural Network (DCNN) automatically extracts additional information from the input image. However, at the initial stage, the input's position and orientation are fully covered. Therefore, this research methodology uses the Kernelized layer with DCNN. In addition, it has a vanishing gradient issue.

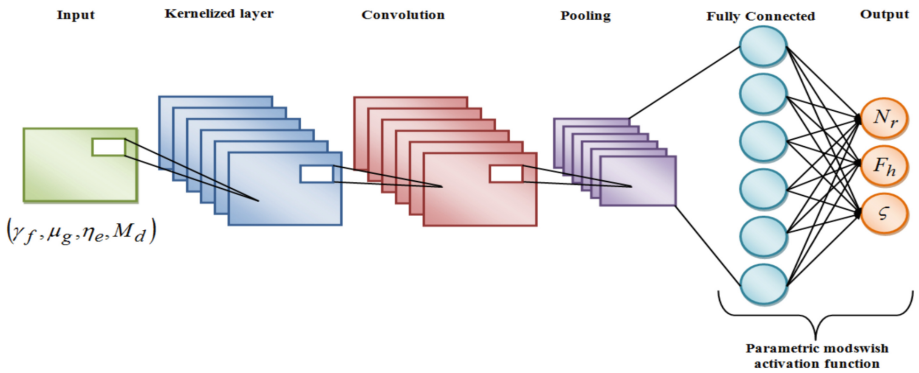


Fig. 2. Architecture of KDCPMNN approach

The Parametric Modswish activation function is utilized in this research methodology to solve the vanishing gradient issue. Figure 2 exhibits the architecture of the KDCPMNN.

Primarily, the combination of γ_f,μ_g, η_e, and M_d is defined as W_j. After that, the position and orientation of W_j are aligned utilizing the Kernelized layer K_{lr}, which is aligned by their inner product calculation.

$$K_{lr} = \xi\left(W_j\right) \cdot \xi\left(W_{j+1}\right) \tag{11}$$

where, the inner product function is indicated as $\xi(.)$. Subsequently, the convolution layer C_{lr} is derived in Eq. (12),

$$C_{lr} = wt.K_{lr} + bs \tag{12}$$

the weight and bias values are notated as wt and bs, respectively. Then, the pooling layer P_{lr} formulation is given in Eq. (13),

$$P_{lr} = \frac{\iota - H(C_{lr})}{S} + 1 \tag{13}$$

$$O_{lr} = \begin{cases} |P_{lr}|, \ if \ P_{lr} > 0 \\ |P_{lr}|.sigmoid \ |P_{lr}| \ else \end{cases} \tag{14}$$

the width is symbolized as ι, the pooling kernel size is signified as $H(C_{lr})$, and the stride is specified as S. After that, the Parametric Modswish activation function is derived in the output layer O_{lr}, which is given in Eq. (14). The loss τ_{lss} is computed in Eq. (15),

$$\tau_{lss} = -\sum \upsilon_{td} \log O_{lr} \tag{15}$$

where, the target value is notated as υ_{td}. The KDCPMNN's pseudocode is given below,

Psuedo code for KDCPMNN

Input: W_j
Output: O_{lr}

Begin
  Initialize K_{lr}, C_{lr}, and τ_{lss}
  For each W_j **do**
    Perform K_{lr}
    Estimate C_{lr}
    Calculate $P_{lr} = \frac{\iota - H(C_{lr})}{S} + 1$
    Calculate O_{lr}
    If $(P_{lr} > 0)$ {
      $O_{lr} = |P_{lr}|$
    } **else** {
      $O_{lr} = |P_{lr}|.sigmoid|P_{lr}|$
    } **end if**
    Calculate τ_{lss}
  End for
  Return O_{lr}
End

The classified output O_{lr} contains normal N_r, fire happens F_h, and potential for present fire ζ.

3.11 Burn Area Assessment

The obtained output is F_h, then the CBI is estimated to assess the burn area. The CBI $\lambda_{cbi}$is derived in Eq. (16),

$$\lambda_{cbi} = \sum f_s / t_n \tag{16}$$

In the Eq. (16) the factor for the obtained class is indicated as f_s, and the total number of factors for the class is signified as t_n. By utilizing the SLM-FIS algorithm, the burn area is assessed based on the calculated λ_{cbi}. A very efficient solution for complex problems is rendered by the conventional Fuzzy Inference System (FIS) algorithm. Nevertheless, the conventional membership function is static and not good for dynamic changes. Therefore, the Secant Linear Membership (SLM) function is utilized in the research methodology.

Initially, the rule G_r is generated in Eq. (17),

$$G_r = \begin{cases} lo_w, if \lambda_{cbi} = [0.25,1.25] \\ mo_d, if \lambda_{cbi} = [1.25,2.25] \\ hi_h, if \lambda_{cbi} = [2.25,3] \end{cases} \tag{17}$$

where, the low, moderate, and high of the affected area are defined as lo_w,mo_d, and hi_h, respectively. After that, the generated rule according to the data is transferred into the crisp data q_{sp}. By using the SLM function, the membership function M_f is constructed based on the transferred data as,

$$M_f = q_{sp} - f(q_{sp}) \frac{f(q_{sp}) - f^0(q_{sp})}{f^1(q_{sp}) - f^0(q_{sp})} \tag{18}$$

Subsequently, the crisp data is transferred into normal data. The SLM-FIS's pseudocode is given below,

Pseudo code for SLM-FIS

Input: F_h
Output: G_r

Begin
  Initialize CBI λ_{cbi}
  For each λ_{cbi} **do**
    If (λ_{cbi} = [0.25,1.25]) {
      $G_r = lo_w G_r = lo_w$
    } **if** (λ_{cbi} = [1.25,2.25]) {
      $G_r = mo_d$
    } **if** (λ_{cbi} = [1.25,2.25]){
      $G_r = hi_h$
    }
    Construct membership function M_f
  End for
End

The abnormal construction design γ_{ab} of the channel is displayed by the SMF-FIS.

3.12 Potential Assessment

The obtained outcome of O_{lr} is ζ, then the confidence score is computed for the factors of obtained classes. It also checks whether the potential assessment is lower-level or higher-level. For this, the confidence score L_{score} is computed utilizing Eq. (19),

$$L_{score} = m \pm con_{fid} \frac{f_{std}}{\sqrt{s_s}} \tag{19}$$

In the Eq. (19) the factor's mean and confidence level value are signified as m and con_{fid}, respectively, the factor's standard deviation is notated as f_{std}, and the factor's size is symbolized as s_s. As per the L_{score}, the potential is assessed utilizing SMF-FIS, which is explained in Sect. 3.11. In this, the rule is generated utilizing Eq. (20),

$$Pot_r = \begin{cases} p_{lo}, & if \quad L_{score} < 25 \\ p_{hg}, & if \quad L_{score} > 90 \end{cases} \tag{20}$$

the generated rule for potential assessment is notated as Pot_r as shown in Eq. (20), the low potential is depicted as p_{lo}, and the high potential is symbolized as p_{hg}.

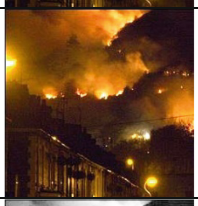
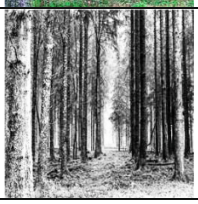
4 Result and Discussion

This section analyzes the performance of the proposed research methodology, which has been implemented using Python as the development platform.

4.1 Dataset Description

The forest meteorological data is gathered from Kaggle for the performance analysis, and the link is given under the reference section. Google Earth map collects Landsat images. From the collected data, 80% of data and 20% of data are used for training purposes and testing purposes, respectively. The sample image output is shown in Table 1.

Table 1. Sample image analysis.

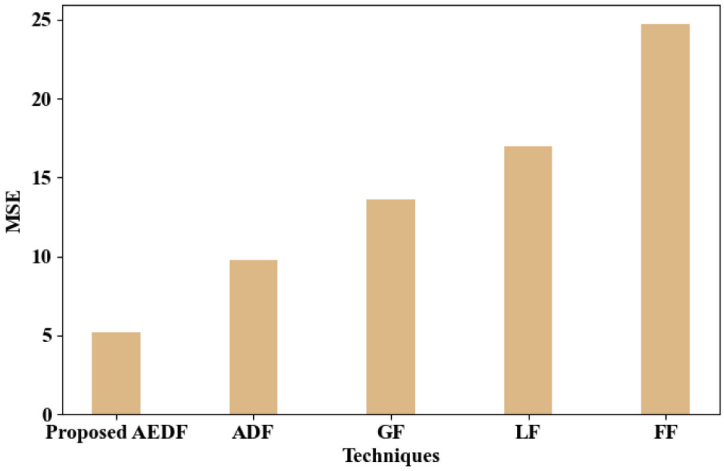
Phases	Sample 1	Sample 2
Input image (after segmentation from satellite image)		
Blurriness removal		
Contrast enhancement		
TM		

4.2 Performance Analysis

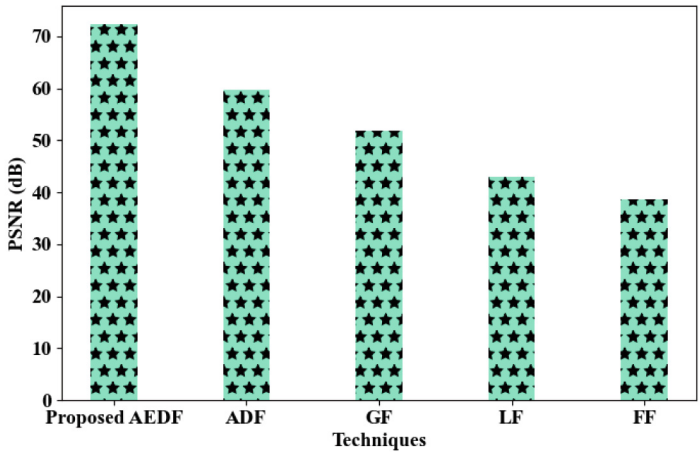
(a) Performance analysis of blurriness removal

In this section, the performance of the AEDF is analyzed with the ADF, Gaussian Filter (GF), Lee Filter (LF), and Frost Filter (FF).

In Fig. 3, the Graphical representation of blurriness removal analysis concerning Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR) is depicted. Owing to the integration of the exponential function, the AEDF's MSE value is low (i.e., 5.2) and PSNR is high (i.e., 72.3 dB).



(a)



(b)

Fig. 3. Graphical analysis of blurriness removal based on (a) MSE and (b) PSNR

(b) Performance analysis for classification

The proposed KDCPMNN's performance is investigated with the DCNN, Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN), and ANN.

The error analysis of the proposed and existing classifiers centered on MSE, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) is shown in Table 2. In this, the KDCPMNN's MSE value is 2.1; however, the conventional techniques have higher errors. Owing to the integration of the kernelized layer with the parametric modswish

Table 2. Error analysis for classifiers

Metrics	Proposed KDCPMNN	DCNN	LSTM	RNN	ANN
MSE	2.1	6	10.4	13.7	18
MAE	4.7	9.5	13.8	16.1	20.5
RMSE	3.4	8.2	12	15.4	19

activation function, the proposed methodology has lower error than the existing research methodologies based on the other metrics also.

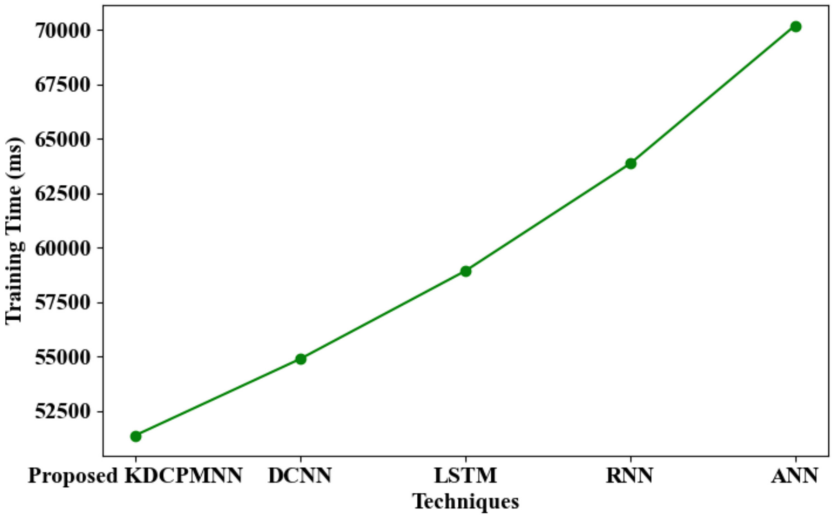


Fig. 4. Training time analysis

In Fig. 4, the training time of the proposed and existing classifiers is exhibited. Here, the KDCPMNN's training time is 51382ms, which is lower compared to the existing research works.

The classifiers' performance is analyzed regarding precision, recall, F-Measure, False Negative Rate (FNR), FPR, and Mathew's Correlation Coefficient (MCC) and is exhibited in Table 3. The F-Measure value of the proposed KDCPMNN approach is higher (97.2%) and the prevailing ANN is lower (87.6%).

The accuracy analysis of the classifiers for FFD is depicted in Fig. 5. The KDCPMNN's accuracy is 98.2%, which is higher than the conventional approaches. The existing research methods' average accuracy is 92.09%, which is lower than the proposed model.

(c) Performance analysis for burning area and potential assessment.

In this, the proposed SLM-FIS's performance is contrasted with the prevailing FIS, Sigmoid Fuzzy Logic (SFC), Trapezoidal Fuzzy Logic (TFL), and Decision Rule (DR).

Table 3. Performance analysis for the classifiers

Metrics	Proposed KDCPMNN	DCNN	LSTM	RNN	ANN
Precision	96.8	94.2	92	90	88.3
Recall	97	93.5	91	89.7	87
F-measure	97.2	94.1	91.5	90.2	87.6
FNR	0.0486	0.0690	0.0903	0.1287	0.1585
FPR	0.0583	0.0796	0.0986	0.1341	0.1682
MCC	97.6	95	93.2	91	89.2

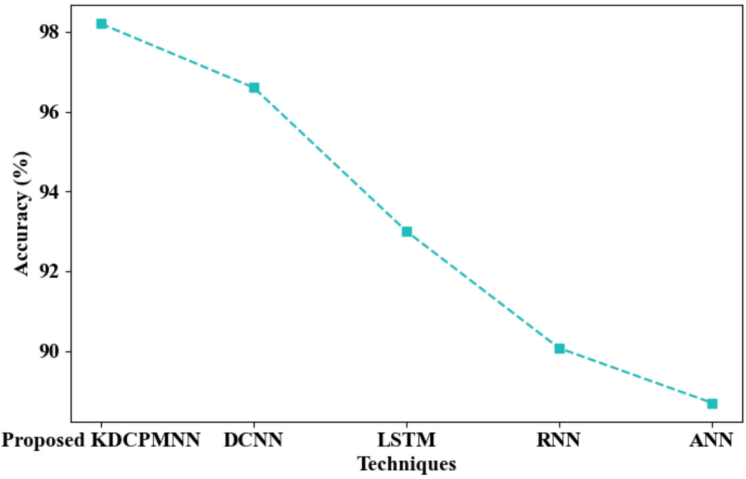


Fig. 5. Accuracy analysis

The performance analysis for the assessment concerning fuzzification, defuzzification, and rule generation time is displayed in Fig. 6. Here, the proposed SLM-FIS has fuzzification time, defuzzification time, and rule generation time of 672ms, 704ms, and 831ms, respectively. Nevertheless, the conventional techniques took more time when analogized to the proposed model.

4.3 Comparative Analysis

In Table 4, the comparative analysis of the proposed and existing research models based on accuracy and F-Measure metrics is exhibited. Here, when compared to other prevailing models, higher accuracy (97.55%) and F-Measure (93.33%) are achieved by the Radial Basis Function Networks with Rapid and Accurate Image Super-Resolution (RBFNs-RAISR). However, RBFNs-RAISR also renders lower performance than the proposed model due to considering meteorological data with advanced techniques.

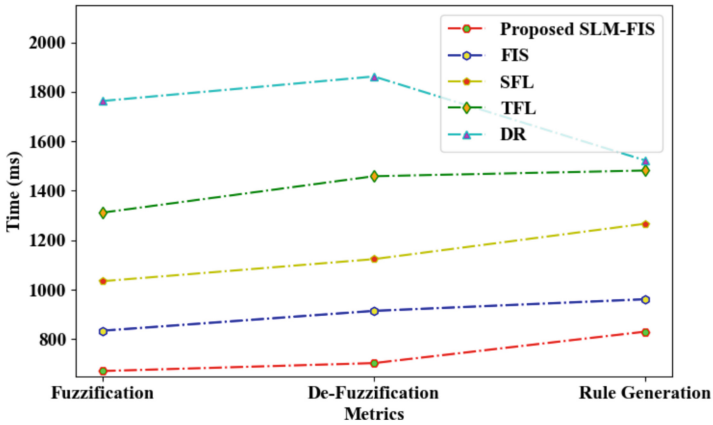


Fig. 6. Performance analysis for the assessment

Table 4. Comparative analysis of proposed and state-of-art works

Author's Name	Methods	Accuracy	F-Measure
(Pang et al., 2022)	RFs	89.2%	-
(Prakash et al., 2023)	Radial basis function Networks with rapid and accurate image super-resolution (RBFNs-RAISR)	97.55%	93.33%
(Pacheco et al., 2021)	k-Nearest Neighbor (kNN) and RF	93%	-
(Rashkovetsky et al., 2021)	U-Net	-	0.83
(Shadrin et al., 2024)	Neural Network	-	0.67
Proposed	KDCPMNN	98.2%	97.2%

5 Conclusion

In this paper, a Kernelized Deep Convolutional Parametric Modswish Neural Network (KDCPMNN) approach is proposed for Forest Fire Detection (FFD) in Sikkim, utilizing remote sensing images obtained from publicly available resources. The results demonstrate an impressive accuracy of 98.2% and a reduced training time of 51,382 ms, indicating the efficiency of the proposed model.Beyond its immediate performance metrics, this study has broader implications for forest fire management and monitoring systems. The high accuracy achieved by the KDCPMNN model suggests its potential applicability in real-time fire detection, allowing for timely responses that can mitigate the impacts of forest fires on biodiversity, ecosystems, and local communities. Additionally, the integration of remote sensing data can enhance the understanding of fire behaviour and spread, providing valuable insights for future research and policy development in forest fire management. Furthermore, this work sets the stage for future studies to explore the scalability of the KDCPMNN approach in different geographical regions

and its adaptability to varying climatic conditions. By establishing a robust framework for forest fire detection and prediction, this research contributes to the growing body of knowledge aimed at developing more effective and proactive forest management strategies. Ultimately, the KDCPMNN approach has the potential to become a crucial tool for environmental monitoring and disaster response, underscoring the importance of leveraging advanced technologies in addressing pressing ecological challenges.

Future Scope: More concentration is rendered for the image, however, the major flaws in meteorological data were not deemed. The research will improve the meteorological data with advanced algorithms in the future.

References

1. Abdusalomov, A.B., Islam, B.M.S., Nasimov, R., Mukhiddinov, M., Whangbo, T.K.: An improved forest fire detection method based on the detectron2 model and a deep learning approach. Sensors **23**(3), 1–19 (2023). https://doi.org/10.3390/s23031512
2. Arisanty, D., Muhaimin, M., Rosadi, D., Saputra, A.N., Hastuti, K.P., Rajiani, I.: Spatiotemporal patterns of burned areas based on the geographic information system for fire risk monitoring. Int. J. Forestry Res. **2021**, 1 (2021). https://doi.org/10.1155/2021/2784474
3. Banerjee, P.: Maximum entropy-based forest fire likelihood mapping: analysing the trends, distribution, and drivers of forest fires in Sikkim Himalaya. Scand. J. For. Res. **36**(4), 275–288 (2021). https://doi.org/10.1080/02827581.2021.1918239
4. Bar, S., Parida, B.R., Pandey, A.C.: Landsat-8 and Sentinel-2 based Forest fire burn area mapping using machine learning algorithms on GEE cloud platform over Uttarakhand, Western Himalaya. Remote Sens. Appl. Soc. Environ. **18**, 1–11 (2020). https://doi.org/10.1016/j.rsase.2020.100324
5. de Almeida Pereira, G.H., Fusioka, A.M., Nassu, B.T., Minetto, R.: Active fire detection in Landsat-8 imagery: a large-scale dataset and a deep-learning study. ISPRS J. Photogramm. Remote Sens. **178**, 171–186 (2021). https://doi.org/10.1016/j.isprsjprs.2021.06.002
6. Di Napoli, M., Marsiglia, P., Di Martire, D., Ramondini, M., Ullo, S.L., Calcaterra, D.: Landslide susceptibility assessment of wildfire burnt areas through earth-observation techniques and a machine learning-based approach. Remote Sens. **12**(15), 1–28 (2020). https://doi.org/10.3390/RS12152505
7. Florath, J., Keller, S.: Supervised machine learning approaches on multispectral remote sensing data for a combined detection of fire and burned area. Remote Sens. **14**(3) (2022). https://doi.org/10.3390/rs14030657
8. Guede-Fernández, F., Martins, L., de Almeida, R.V., Gamboa, H., Vieira, P.: A deep learning based object identification system for forest fire detection. Fire **4**(4), 1–17 (2021). https://doi.org/10.3390/fire4040075
9. Hang, H.T., Mallick, J., Alqadhi, S., Bindajam, A.A., Abdo, H.G.: Exploring forest fire susceptibility and management strategies in Western Himalaya: Integrating ensemble machine learning and explainable AI for accurate prediction and comprehensive analysis. Environ. Technol. Innov. **35**, 1–23 (2024). https://doi.org/10.1016/j.eti.2024.103655
10. Hernandez-Lopez, D., Lopez-Rebollo, J., Moreno, M.A., Gonzalez-Aguilera, D.: Automatic processing for identification of forest fire risk areas along high-voltage power lines using coarse-to-fine LiDAR Data. Forests **14**(4), 1–20 (2023). https://doi.org/10.3390/f14040662
11. Hong, Z., et al.: Active fire detection using a novel convolutional neural network based on Himawari-8 satellite images. Front. Environ. Sci. **10**, 1–11 (2022). https://doi.org/10.3389/fenvs.2022.79402

12. Kang, Y., Jang, E., Im, J., Kwon, C.: A deep learning model using geostationary satellite data for forest fire detection with reduced detection latency. GIScience Remote Sens. **59**(1), 2019–2035 (2022). https://doi.org/10.1080/15481603.2022.2143872
13. Khan, A., Hassan, B., Khan, S., Ahmed, R., Abuassba, A.: DeepFire: a novel dataset and deep transfer learning benchmark for forest fire detection. Mob. Inf. Syst. **2022**, 1–14 (2022). https://doi.org/10.1155/2022/5358359
14. Khryashchev, V., Larionov, R.: Wildfire segmentation on satellite images using deep learning. In: Moscow Workshop on Electronic and Networking Technologies, MWENT 2020 - Proceedings, pp. 1–5 (2020). https://doi.org/10.1109/MWENT47943.2020.9067475
15. Pacheco, A.D.P., Junior, J.A.D.S., Ruiz-Armenteros, A.M., Henriques, R.F.F.: Assessment of k-nearest neighbor and random forest classifiers for mapping forest fire areas in central Portugal using landsat-8, sentinel-2, and terra imagery. Remote Sens. **13**(7), 1–25 (2021). https://doi.org/10.3390/rs13071345
16. Pang, Y., et al.: Forest fire occurrence prediction in China based on machine learning methods. Remote Sens. **14**(21), 1–23 (2022). https://doi.org/10.3390/rs14215546
17. Pinto, M.M., Libonati, R., Trigo, R.M., Trigo, I.F., DaCamara, C.C.: A deep learning approach for mapping and dating burned areas using temporal sequences of satellite images. ISPRS J. Photogramm. Remote Sens. **160**, 260–274 (2020). https://doi.org/10.1016/j.isprsjprs.2019.12.014
18. Prakash, M., Neelakandan, S., Tamilselvi, M., Velmurugan, S., Baghavathi Priya, S., Ofori Martinson, E.: Deep learning-based wildfire image detection and classification systems for controlling biomass. Int. J. Intell. Syst. **2023**, 1–18 (2023). https://doi.org/10.1155/2023/7939516
19. Rashkovetsky, D., Mauracher, F., Langer, M., Schmitt, M.: Wildfire detection from multisensor satellite imagery using deep semantic segmentation. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. **14**, 7001–7016 (2021). https://doi.org/10.1109/JSTARS.2021.3093625
20. Sathishkumar, V.E., Cho, J., Subramanian, M., Naren, O.S.: Forest fire and smoke detection using deep learning-based learning without forgetting. Fire Ecol. **19**(1), 1–17 (2023). https://doi.org/10.1186/s42408-022-00165-0
21. Shadrin, D., et al.: Wildfire spreading prediction using multimodal data and deep neural network approach. Sci. Rep. **14**(1), 1–17 (2024). https://doi.org/10.1038/s41598-024-52821-x
22. Wahyono, H.A., Dharmawan, A., Adhinata, F.D., Kosala, G., Jo, KH.: Real-time forest fire detection framework based on artificial intelligence using color probability model and motion feature analysis. Fire **5**(1), 1–15 (2022). https://doi.org/10.3390/fire5010023
23. Xu, R., Lin, H., Lu, K., Cao, L., Liu, Y.: A forest fire detection system based on ensemble learning. Forests **12**(2), 1–17 (2021). https://doi.org/10.3390/f12020217

Decentralized Payment Gateway: Bridging DeFi with Metamask

Ashish Revar(✉), Khooshi Sonkar, and Kalpesh Wandra

SITAICS, Rashtriya Raksha University, Gujarat, India
{ashish.revar,kalpesh.wandra}@rru.ac.in,
20bcscs056@student.rru.ac.in

Abstract. When online transactions are involved, the act of transferring money most of the time necessitates the sharing of personal details such as names, bank accounts, and credit card numbers. The main mode of the transactions is through centralized systems, and they require a long period to settle. Emerging technologies innovate the market which necessitates a fast and secure payment system. The DeFi model relies on blockchain focusing on the distributive approach as its main tool. Blockchain is a digital application of allocation transactions by a decentralized system that is used for its security and transparency. DeFi employs the technology of blockchain and smart contracts to have financial services dispersed in a decentralized manner. This paper accepts a decentralized payment system utilizing Distributed Finance (DeFi) and blockchain. The system is composed of a frontend website and backend smart contracts, which perform the functions of user interface payment processing, and ledger updates. It integrates MetaMask, which is a cryptocurrency wallet, that stores user accounts and cryptographic keys, and also interacts with the website. Utilizing this approach, the site users can make the sending and receiving of payments super fast without the knowledge of the fact that the website interacts with the blockchain in the background through MetaMask. There's no need to wait for hours in line to have your transaction settled, and you don't have to disclose personal details. Blockchain does not have just one single node that controls everything because of the decentralized plan of the system, it is transparent, secure, and efficient (accessible). With this knowledge, this paper shows how comfortable DeFi and blockchain technologies are in developing better payment forms.

Keywords: Decentralized Payment Gateway · MetaMask Integration · DeFi (Decentralized Finance) · Smart Contracts · Ethereum Blockchain

1 Introduction

Since the enhancement of the technologies over the internet, financial transactions are carried miles away online. However conventional payment solutions are characterized by weak protection of users' private data, the implication of middlemen, and, therefore, risks of personal data leaks, high costs, and extended time to finalize transactions. As technology advances, there are increasing need for advanced, efficient, secure, and

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 395–408, 2025.
https://doi.org/10.1007/978-3-031-86305-9_27

cheaper methods of transacting. Considering the subject, Blockchain has risen to the position of being a revolutionary in the financial world whereby it provides an open platform for recording all or some of the transactions. In simple terms, it is an electronic ledger, which works as a database that is shared by nodes in the structure of a chain of blocks to store transactions. This structure makes it less centralized with accountability recorded permanently so it cannot be manipulated. DeFi is an emerging financial system to rebuild centralized financial structures into a de-centralized architecture; away from the control of conventional financial intermediation. Key components of DeFi include:

1. Smart Contracts: It is defined as a self-automated contract where the terms of the contract are written in the form of functions.
2. Decentralized Applications (DApps): Specialized applications which operate in a network when connected between different computers instead of within a single computer.
3. Decentralized Exchanges (DEXs): In this case, it provides a platform where users can transact with other users directly using cryptocurrency without the involvement of any outsider.

Our proposed payment gateway is based on the decentralized finance principles where overall control is moved from a central point to distributed nodes, which makes it faster, more secure, and easier to use. This strategy will integrate blockchain with familiar web interfaces, so we enable the move from simple web payments to a new world of Web3 and DeFi. Here in this paper, the design and general deployment of this decentralized payment gateway are highlighted, together with some envisaged prospects. Let us examine how this system can overcome some of the challenges that face current online payment processing including, slow processing duration, expense per transaction cost, and user privacy. Additionally, we explain how we incorporate MetaMask, an Ethereum-supported cryptocurrency wallet, to enhance the usability of the system and preserve the properties of blockchain-based security. Concerning the details of this system and its possible implications, we hope that this paper will make a positive contribution to debates regarding the prospects of financial technologies in the development of a fair, effective, and safe global financial environment.

2 Related Work

A decentralized payment gateway enables direct peer-to-peer economic transactions without third-party intermediaries using blockchain technology integration. Ethereum public blockchain infrastructures offer inherent capabilities like transparency, immutability, security, and elimination of single points of failure that are the main problems of traditional fintech systems. In this situation, the Ethereum-based distributed digital platform works as a virtual machine where a smart contract [1].

The problem of expensive fees, delayed settlement time, and opaque governance models in the existing payment networks will be tackled by the integration of digital wallets such as MetaMask [2]. Our proposed gateway architecture contains a ReactJS frontend interface as well as an Ethereum-based backend for transaction handling.

Some blockchain-integrated payment solutions use React.js to build web interfaces with the user in mind. The language is declarative, and it is open source; besides, its

executions are predictable and easy to debug. Here the react hooks assist in helping in manipulation of the DOM [3]. React with a declarative programming model does make the whole development a lot easier compared to jQuery which uses the imperative model. React's responsive design takes into consideration the availability across the devices. As a whole, React gives an extendable and scalable frontend structure to deal with the decentralized application's requirements.

Web3.js connectors configure the client-side wallet connections, while smart contracts written in Solidity enforce the back-end rules related to escrow, transfers, and immutable storage [4]. Stablecoins through DeFi protocols or direct fiat transfers via Synapse bridge the fiat-to-crypto world to provide flexible settlement options. Permissionless innovation will be facilitated by continuous governing based on DAO structures.

A system architecture of a blockchain payment system which consists of a React frontend, Ethereum smart contracts for transaction logic, and MetaMask for wallet connectivity is proposed. The modular design enabled the layering of the blockchain complexities into the user experience. The use of smart contracts enabled real-time transfer of money between the parties. MetaMask helped to approve the transactions by users. Still, the blockchain scalability problem was being outlined as some performance issues were encountered. The use of blockchain improves the flow of information in the supply chain and transparency that various firms can adapt to trace their commodities from the manufacturing firm to the consumers [5]. This will be useful in checking fraud and other unlawful deeds, and therefore increase efficiency in the supply chain. Though the system made it easy already for cross-border payments, banking system integration for the final settlement was still necessary.

DeFi is a monetary system operating on decentralized public blockchain networks [6]. They achieved this by utilizing a decentralized payment gateway using technology DeFi protocols on Ethereum. They designed a React front-end that interacted with smart contracts that enabled transaction control and ledger updating. Integration of MetaMask allowed the users to easily manage accounts and sign transactions digitally. This technology eliminated the role of traditional intermediaries and made the web-based payment system to be familiar. On the other hand, the DeFi-based payment gateway had a simple way of combining decentralized payments with ecommerce and other use cases. It identified the existing research challenges such as UX improvements, fiat on and off-ramps, and regulatory considerations that must be fulfilled for the mainstream adoption of decentralized finance.

Most of the e-commerce website use this traditional payment methods to make their user to make digital payments, assuming that users will make digital payments using their bank balances or credit cards [7]. This has motivated us to develop a prototype e-commerce website that uses decentralized technology, aiming to enhance security and overcome the digital payment fraud for users. We propose to develop an e-commerce website using React and integrating Metamask, from where the users can pay straight from their Ethereum wallets. For instance, the MetaMask enables users to shop via the web app. Thus, users receive the desired product with which they can complete a purchase using a wallet application already known to them. The smart contracts took control of processing disposal orders as well as inventory management [7]. In contrast to the

established e-commerce systems, the blockchain system brought about quick settlement of transactions, low fees, and greater privacy. Nevertheless, e-commerce with decentralized payments has disruptive potential because it eliminates payment intermediaries and decreases fees. Security is the main issue behind designing of Gateway. That's why the proposed architecture is made by the implementation of secure electronic transaction methods [8].

3 System Architecture

3.1 User Friendly Website

This web page makes the sending and receiving of money by the folks to be quite easy. It would be available from lots of mobile phones, computers, tablets – any device that you can have access to your web browser. The platform puts onto the screen an amount of money that is now in the user's account and all the earlier transactions. Several forms to open them are there which keep tabs on the amount as well as on the payee of the transaction.

3.2 Smart Contracts

Smart contracts are protocols validated within the blockchain environment and can be referred as functions embedded in the blockchain to initiate and execute any transaction [9]. They are involved in the transaction verification and packages of data (payment logic). When sending the cash the website emails the contract and also changes the money transfer. The contract would check the balances and then transfer the amount if the transaction would be real. The number of recordings is very large, as the blockchain is updated to record these many payments.

3.3 MetaMask Wallet

This digital wallet plugin acts as the link between the website and the blockchain right in the middle. The middleman between the website and the blockchain is this digital wallet plugin. Wallet in this context is a blockchain wallet and this type of wallet includes an address or Public Key (PK) and Secret Key (SK) [10]. They download the add-on and add it to their browser. It has their private keys safely. A user is prompted by MetaMask to approve the transaction when he/she is undertaking a payment. Besides, as these contract transactions are being executed in real-time, one can verify proof of these transactions on Metamask and Etherscan [11]. After that, the nonce is appended, and the transaction is signed by the e-wallet on behalf of the user using cryptography. Not only it provides necessary storage and key management but can also connect with hardware-based crypto wallets, exchange cryptocurrencies, interact with numerous smart contracts, etc. [12].

Process Overview:

1. The user goes to the payment website and links his/her MetaMask account.

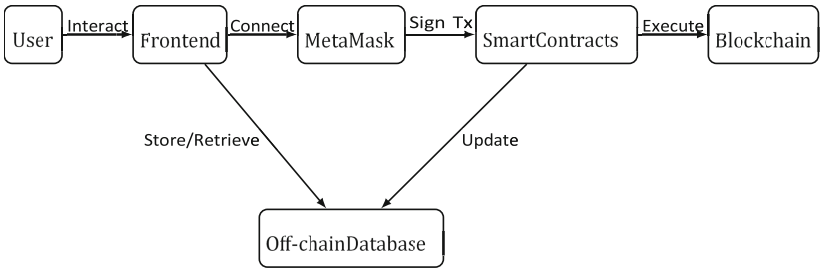


Fig. 1. System Architecture

2. User inputs all the relevant information regarding the money transfer. The website starts the payment process.
3. MetaMask prompt comes out for the user to confirm the payment. The user approves.
4. MetaMask communicates with and allows lots of websites to send funds to smart contracts.
5. Blockchain automatically keeps track of the transfer.

In conclusion, the system we propose utilizes the website as the user interface, smart contracts to implement the payment logic, and also MetaMask to integrate with the blockchain. In unison, they facilitate smooth, secure payments on the internet and also in banking.

4 Frontend Implementation

The user interface is the place where the customers make their purchases. It should function as expected for the user of a money transfer app. ReactJS will be used in building a lot of the website. React JS is a very highly popular framework for the development of web and also mobile applications. It creates reusable items in its user interface.

Frontend design:

- Home page - there, the users are met and there is a login/register button.
- Dashboard - The account balance and activity can be reviewed here.
- Send money - Indicate the amount and the person.
- Receive Money: Shows the wallet address to which you should send the money.
- Action - Refers to the previous transfers and also the entries of the transactions.

The design optimizes for rapid execution and clearing, with slower settlement [13]. The goal will be accompanied by many explanations and illustrations that will be very useful for the users. When the users are transferring the money, they will type the amount and the recipient on the send money page. From there, they will go ahead and click on Submit to initiate the money transfer. In the stage below, React is required to connect with the smart contract of the public blockchain through the Web3 code. This is safe

as the funds are being forwarded and the ledger is being updated. So React has to wait for a blockchain confirmation and shows a pending status. When status completed. The customer interface is being redesigned with the new balances. The users are always aware of the incoming payments through the notifications they receive. Consequently, in short, ReactJS will present an up-to-date and usable interface for such cases of daily international money transfers. Powered underneath is our secure smart contract backend connected to the blockchain ready to collect and release the payments (Figs. 1 and 2).

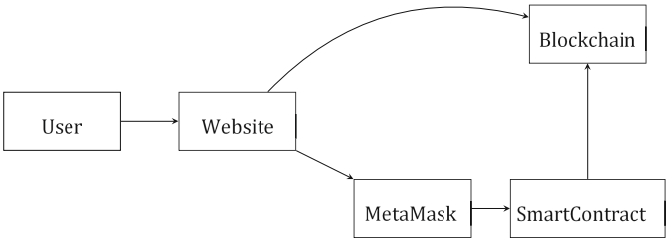


Fig. 2. Frontend Processing

5 Smart Contract Implementation

Smart contracts are contracts which automatically trigger the execution of a set of terms codified on them. It uses Ethereum and Solidity by which it triggers automatic actions based on certain conditions. They work as the primary engine of the payment gateway. In this paper, we have proposed to develop smart contracts using Solidity which is the language of Ethereum blockchain Dapps.

```
pragma solidity ^0.8.0;
contract Payment { mapping(address => uint) public bal-
ances;
function deposit() public payable { balances[msg.sender]
+= msg.value;
}
function transfer(address payable recipient, uint amount)
public { require(balances[msg.sender] >= amount, "Insuf-
ficient balance"); balances[msg.sender] -= amount; recip-
ient.transfer(amount);
}
}
```

This solidity code consists represents a basic payment contract in which users can store Ether and transfer amount to others. Deposit and transfer are two elements available in the application; Deposit function—a method of adding funds to the user's balance, Transfer function—a method of forwarding funds to a recipient with a proper balance check before sending money. Including such examples will help the reader who may

be unaware of smart contracts or want to get additional knowledge about this concept (Fig. 3).

The core smart contracts will handle:

- User Accounts – Balancing of Amounts
- Transfer function - Logic of payment between users.
- Ledger updates - Adding the transactions to the chain

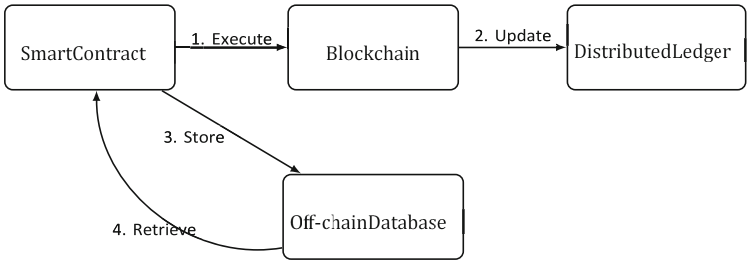


Fig. 3. Backend Processing

Ether is the other cryptocurrency of Ethereum primarily employed in transaction fees. The accounts of two types: those are the external ones, the encrypted ones with their private keys, and the contract accounts resulting from messages that initiate the execution of the predefined contract conditions [22]. When the front end of the website is developed to the stage where it is ready to move the funds it will utilize the transfer function of the contract. Recipients of payment orders are to give descriptions of the payment to the sender, and the contract provides that the sender should have sufficient money. It will curtail the amount from the sender and add the amount to the receiver balance in the amount that has been transferred. It will subtract from the sender balance and add to the receiver balance by the transfer amount. At the last stage of the contract, the ledger on the blockchain is amended with the transaction details. This performs the payment. We will undertake to conduct the tests and audits of the security contracts. Consequently, Solidity smart contracts will allow the ground payment logic on blockchain platforms to work safely and efficiently.

6 Metamask Integration

MetaMask which helps in connecting the Ethereum accounts and safely engage with the blockchain, is used here as the browser extension wallet [14]. Incorporate it into our payment system for a better user experience. Both users should do a MetaMask browser extension download and installation on their computer or mobile device. This gives them a fully protected cryptocurrency wallet where they can store their money. Customers will see that our payment website operates with the use of blockchain when they visit it and

it will be noticed by MetaMask. It will prompt the user to plug in the wallet. If it is the user's need, they are connected. We use the Web3 library to communicate with our frontend Metamask. Web3 shoulders all the hard and tedious processing of backend blockchains.

Here is how users send money with MetaMask:

1. The customer enters payment details on the website.
2. The website deals with transfer functions of smart contracts.
3. MetaMask pops the question for user permission as regards the transaction coverage.
4. Therefore, if the user approves, MetaMask generates a cryptographic signature.
5. Signature is the sign-off value for payment authorized by the user.
6. Web3 signs the contract by sending the signatures to the specified contract.
7. Contract makes up for the transfer, update, and maintenance of records on the blockchain.
8. Website confirms that the user's money has been transmitted.

And hence, in short, MetaMask gives the users ownership of the e-wallets in the shape of blockchain wallets. With the help of MetaMask, web3.js has got front-end integration. This helps users to confirm transactions fast resulting in the fulfillment of no-doubt smart contract actions. This therefore enables a frictionless decentralized payment experience. Thus, Solidity specifically addresses the issue of the basic payment logic on the blockchain platforms which needs to be safe and working in a performant way (Fig. 4).

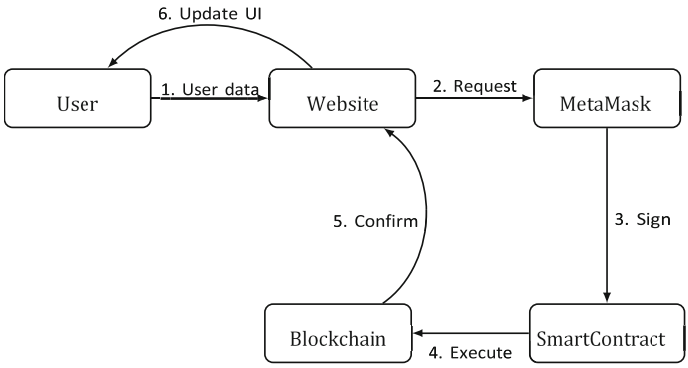


Fig. 4. Illustration of how MetaMask interacts with smart contracts

7 Proposed Empirical Findings and Performance Analysis

Since we are still in the idea stage for our decentralized payment gateway, the following figures have been derived from previous DeFi platforms and existing blockchain technology. These projections are intended to guide implementation and highlight our systems expected benefits.

7.1 Estimated System Throughput

Taking into account the current Ethereum and the Layer 2 roll-up networks solutions, we can calculate the transaction output with this below equation:

$$T = \min\left(\frac{N}{B}, C\right) \tag{1}$$

where,

- N is the number of transactions in a block.
- B is the average time of creation of block in descending order in terms of seconds.
- C represents the maximum number of transactions per seconds over the network.

For ethereum, coefficient B is approx. Equal to 15 s [15]. Assuming N = 200 transactions per block and C = 20 transactions per second, we get:

$$\mathrm{T} = \min\left(\frac{200}{15}, 20\right) = 13.33\mathrm{transactionspersecond} \tag{2}$$

This inline that peak performance takes up to 15–20 transactions per second, wherein sustained performance takes up to 10–15 transactions per second.

This equation also expresses the limitations of blockchain in terms of transaction processing. It is limited to the block capacity (N/B) or the processing capacity of the network (C) whichever is less. The values used are standard for Ethereum so it can be seen that the projected performance of our system is feasible within the limit of blockchain technology today [16].

7.2 Projected Transaction Speed

Given that a typical time for an Ethereum block is about 15 s [15], we can calculate the transaction output (T) of our system using a simple equation:

where, K is the number of blocks needed for confirmation of transaction for security.

$$E = B \times K \tag{3}$$

Suppose K = 6 for high-security transactions [17]:

$$E = 15 \times 6 = 90 seconds \tag{4}$$

This supports our initial estimation that a full transaction can take between 3–5 min for confirmation.

When using blockchains, more than one confirmation is needed to close the transaction. Thus each confirmation marks a new block which is added after the block of the particular transaction, making the process difficult for reversing the transaction. There are specific confirmations for various basic and advanced blockchains, and the requirement for six confirmations is quite common in many blockchain applications, particularly for a high-value transactions [18].

7.3 Projected Transaction Costs

These costs result from a breakdown of gas costs for similar DeFi applications:

$$C = G \times P \times E \tag{5}$$

where:

- G is the gas that is utilized by the transaction
- P is the gas price (in ETH)
- E as the rate of exchange from ETH to USD

For a typical token transfer, (G $\approx$ 21,000 gas) with P = 20 Gwei and E = 2000USD/ETH:

$$C = 21000 \times \left(20 \times 10^{-9}\right) \times 2000 = 0.84USD \tag{6}$$

This is within our projected figures of 0.10 USD - 2.00 USD per transaction. It reflects the actual cost determination in Ethereum transactions. 'Gas' is the computation price of processing computations where each operation is charged by the agreed amount of gas. Currently, the gas price depends on network traffic and the total cost in fiat money depends on the current rate of ETH [19].

These projections are based on the current status of blockchain technology, and they might change for the better with new improvements in Ethereum or with the help of other kinds of blockchain technologies available in the market.

8 Applications

It is possible that our decentralized crypto payment gateway can serve as the payment solution in lots of real-world scenarios.

1 **E-Commerce** - The clients could order goods from the shops online without their credit card details. Blockchain allows for immediate transfers without the involvement of banks.
2 **Money transfers around the world** – Foreign workers might see the family members' money at home pretty soon and also it's cheap. In reality, there are no such payments or invoices.
3 **Peer-to-peer payment** - The contacts would no longer need any third party, such as the end of ordering a meal or paying the subscription to a sports channel, no more using Paypal.
4 **Business Invoices** - The firm charges their client upon completion of the work.

The blockchain assures speedy settlement.

In addition to payments, the blockchain data helps with:

- All details are mentioned in the reporting.
- The regulators can also audit payments at the same time they do the compliance audit.
- Analytics - The firms analyze customer purchase data.

To be precise, this implies that the decentralized payment gateway provides tremendous opportunities because the limitless scope makes you the only one who can think of endless possibilities. The vast places where fast, cheap, and transparent payment systems are much needed, blockchain can be really helpful.

9 Limitations and Future Work

While our proposed decentralised payment gateway has substantial benefits, it is critical to address the limitations and concerns that should be considered throughout execution.

9.1 Scalability Parameters

1. Ethereum Network Congestion: when users are highly active, the network on which our system is developed (Ethereum) tends to become congested. This could lead to an increase in transaction delays and a rise in gas prices [20]. Future developments should include:

- Adoption of layer 2 scaling solutions for blockchain web applications, for instance, an optimistic rollup or a zk-rollup [21].
- Looking into more innovative and flexible blockchain technologies such as Solana or Algorand.

2. Transaction Throughput: However, as far as the current state of TPS payment systems and the technology proposed in blockchain, such as in numbers cannot impress the general centralized [22]. It's important to continuously improve this technology to ensure it meets the demands of the market and provides real value to users. Suggestions include:

- Creating state channel networks dedicated to micro-transactions.
- Utilizing sharding techniques to improve the efficiency of the parallel processing of transactions [23]

9.2 Security Enhancements

1. Smart Contract Vulnerabilities: Nevertheless, smart contracts are still susceptible to additional vulnerabilities that ports it states cannot foresee despite thorough tests. Developers ought to:

- Use formal methods as demonstrated by Bhargavan et al. [24].
- Integrate the patterns of upgradeable smart contracts to make provisions for security updates.

2. MetaMask Security: the metamask software is operable through a browser and thus offers convenience but remains weak to phishing attempts. Further studies should include:

- Incorporating phishing attack prevention measures.
- Educating users about best practices for wallet security inadvance.

3. Private Key Management: users have the obligation of protecting the private keys which they own. In order to reduce the risks involved:

Look for possible embedding within hardware wallet solutions.

- Implement social recovery mechanisms in case of lost keys.

9.3 User Adoptions Strategies

1. Technical Complexity: the system requires users to understand concepts like wallets and gas fees. To address this:
- Create the user interfaces which can hide the complexity of the various blockchain operations.
- Put in place detailed user onboarding processes.

2. Volatility Management: cryptocurrency price volatility can impact the way the payment system is viewed in terms of its stability. Existing solutions Include:

- Using stablecoins which are integrated into the protocol.
- Implementing automatic currency conversion features.

Tackling these limitations and challenges will be important in ensuring that the proposed decentralized payment gateway is adopted widely and operates successfully in the long run. We are looking into these aspects in the course of the implementation of the said provisions and in the course of the further development of the system with an aim of making it more effective and more user-friendly in integration of traditional financing and the new-age DeFi systems.

10 Conclusion

This paper sets out to describe a new decentralized payment system with a basis in blockchain technology. A valid alternative to the traditional centralized approaches, it shall sort all those problems that online payments face, regarding speed, cost, and privacy, thanks to Distributed Finance and blockchain.

Some key contributions of our work:

- Simple frontend application that makes transferring and receiving money without any intermediaries a hassle.
- Use of the blockchain will allow smart contracts to securely and efficiently perform the transaction and handle the changes in ledger simultaneously.
- Authentication of users will be carried out by MetaMask wallets in an effortless, intuitive manner, and this wallet will do transaction signing.
- The front end, smart contracts, and MetaMask integration is all non-disruptive and hassle-free to deliver seamless experience for the users.
- Looking forward, we see the following future research and development:
- Cross-chain Compatibility: Extending the system to support multiple blockchain networks, might improve the ability to be more diverse and user-friendly.

- Decentralized Identity Solutions: Including decentralized identity protocols can make it easy for the user to onboard while ensuring respect for all privacy and security requirements.

In conclusion, our decentralized payment gateway is a necessary step toward more efficient, secure, and accessible online transactions. Systems like ours possess the potential to restructure the global and sustain financial inclusion by gradually reducing the traditional intermediaries with the use of DeFi. While significant barrier do exist, the potential of this technology is enormous, and upcoming findings and development in this field will certainly inform and drive some of the future of digital payments and more complete financial systems.

Compliance with Ethical Standards.

Conflict of Interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethical Approval. This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Sholeh, M., Talahaturuson, E.Y., Rizqi, M., Gumelar, A.B.: Designing an ethereum-based blockchain for tuition payment system using smart contract service. Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi) (2022)
2. Chen, Y., Bellavitis, C.: Blockchain disruption and decentralized finance: the rise of decentralized business models. J. Bus. Ventur. Insights **13**, 00151 (2020)
3. Brindashree, C.B.: Implementation of canteen automation system with payment gateway using MERN stack. Int. J. Res. Appl. Sci. Eng. Technol. (2021)
4. Bodkhe, S.P.K.K.P.T.S., Tanwar, U., Kumar, N., Alazab, M.: Blockchain for industry 4.0: A comprehensive review. IEEE Access **8**, 79764–79800 (2020)
5. Srivastava, A.: Design and implementation of a blockchain-based payment solution. Int. J. Res. Appl. Sci. Eng. Technol. (2023)
6. Vishnu, S.: De-fi based payment gateway system. Int. J. Res. Appl. Sci. Eng. Technol. (2023)
7. Aksoy, B.Y., Abri, R.: Exploring the potential of decentralized currency for ecommerce: case study of an e-commerce website using decentralized currency. Int. Conf. Pioneer Innov. Stud. (2023)
8. Pande, A.K., Deshmukh, A.B.: E-payment gateway model (2014). https://api.semanticscholar.org/CorpusID:14195543
9. Christidis, K., Devetsikiotis, M.: Blockchains and smart contracts for the internet of things. Ieee Access **4**, 2292–2303 (2016)
10. Pentinsaari, J.: Private key vulnerabilities in browser wallets (2023)
11. Khan, M., Imtiaz, J., Islam, M.N.U.: A blockchain-based secure decentralized transaction system for energy trading in microgrids. IEEE Access (2023)
12. Kersic, V., Vidovic, U., Vrecko, A., Domajnko, M., Turkanovic, M.: Orchestrating digital wallets for on-and off-chain decentralized identity management. IEEE Access (2023)
13. Poon, J., Team, O.: Decentralized exchange and payments platform (2017)
14. Nair, R.R.: Blockchain-enabled supply chain: a review of decentralized application (DAPP) implementations (2023)

15. Wood, G., et al.: Ethereum: a secure decentralised generalised transaction ledger. Ethereum Project Yellow Paper **151**(2014), 1–32 (2014)
16. Park, D., Zhang, Y., Rosu, G.: End-to-end formal verification of ethereum 2.0 deposit smart contract. In: Computer Aided Verification: 32nd International Conference, CAV 2020, Los Angeles, CA, USA, July 21–24, 2020, Proceedings, Part I 32, pp. 151–164. Springer (2020)
17. Nakamoto, S., Bitcoin, A.: A peer-to-peer electronic cash system. Bitcoin **4**(2), 15 (2008). https://bitcoin.org/bitcoin.pdf
18. Antonopoulos, A.M., Harding, D.A.: Mastering Bitcoin. " O'Reilly Media, Inc.", ??? (2023)
19. Buterin, V., et al.: A next-generation smart contract and decentralized application platform. white paper **3**(37), 2–1 (2014)
20. Badari, A., Chaudhury, A.: An overview of bitcoin and ethereum white-papers, forks, and prices. Forks, and Prices (April 26, 2021) (2021)
21. Gudgeon, L., Moreno-Sanchez, P., Roos, S., McCorry, P., Gervais, A.: SoK: Layertwo blockchain protocols. In: Financial Cryptography and Data Security: 24th International Conference, FC 2020, Kota Kinabalu, Malaysia, February 10–14, 2020 Revised Selected Papers 24, pp. 201–226. Springer (2020)
22. Croman, K., et al.: On scaling decentralized blockchains: (a position paper). In: International Conference on Financial Cryptography and Data Security, pp. 106–125. Springer (2016)
23. Wang, J., Wang, H.: Monoxide: scale out blockchains with asynchronous consensus zones. In: 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI 19), pp. 95–112 (2019)
24. Bhargavan, K., et al.: Formal verification of smart contracts: short paper. In: Proceedings of the 2016 ACM Workshop on Programming Languages and Analysis for Security, pp. 91–96 (2016)

Impact of Social Media Marketing in the Cosmetic Industry to Cross the Chasm in India

Narendra Rathnaraj(✉), C. Rajasree, S. Amrutha, and Arul Rajan

PSG Institute of Management, PSG College of Technology, Coimbatore, Tamil Nadu, India
naren@psgim.ac.in

Abstract. The Indian cosmetic industry is developing at a rapid pace with improvements in digital technologies, which in turn has encouraged the growing number of smaller brands to use social media marketing activities as a significant means toward market success. This research investigates how social media marketing activities (SMMA) influences cosmetic brands in their capability of bridging the gap between early adopters and mainstream consumers, termed "crossing the chasm". Based on a sample of 385 respondents residing at Coimbatore city, the data has been analyzed regarding the influence of social media sites like Facebook, Instagram, Twitter, and YouTube on brand awareness, customer relationship, and enhancement of social image. By understanding this study, one may appreciate how social media greatly affects the creation of proper brand awareness by increasing customer loyalty and enhancing a favorable social im-age, which are both important in the transformation from niche to broader markets. Factors that are key and interconnected will also be outlined in the main body of findings to provide practical insights for cosmetic brands to extend their market influence through effective social media marketing strategies. The study employs PLS – SEM to under-stand he interconnectedness of the variables used in the study. From the analysis it is observed that SMMA has a concrete impact on the marketing of cosmetics in India.

Keywords: Cosmetic industry · Crossing the chasm · Customer Relations · Social media marketing

1 Introduction

The Indian cosmetic industry has witnessed a phenomenal growth, and coupled with mushrooming digital platforms, the manner of marketing and consumption of cosmetic products has dramatically changed. Large, well-known brands that once dominated the sector with expansive distribution networks have, with the assistance of social media, given way to smaller, nimbler businesses that can build consumer relations easily. It becomes more important when considering the aspect of "crossing the chasm", which business consultant Geoffrey Moore has made famous to show how tough the transfer was from early adopters to the mainstream market. The social media marketing activities

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 409–424, 2025.
https://doi.org/10.1007/978-3-031-86305-9_28

(SMMA) have become an important strategy for many cosmetic companies, especially those new in the market, for going beyond this barrier and attracting more consumers. During the digital age, it has become the perfect avenue for brands to reach out and engage with consumers through social networking sites like Facebook, Instagram, and YouTube. These present novel opportunities for companies in the quest for relationship development, community involvement, and content creation that would appeal to both mainstream and early adopters. The visual nature of their format is best suited to the cosmetics industry, in which the buying decisions are highly dependent on how the merchandise looks or appears. Marketers have, therefore, been able to bring into sharp light their products and offer stories in the most engaging ways that catch customers' attention and create loyalty through high-quality pictures, guides, and even social media influencer endorsements.

The change in consumer preferences and the changing market trends in the cosmetic business, facilitate the examination of the impacts of social media marketing. The increase in disposable income in the growing middle-class segment of the population has increased the demand for beauty and personal care products, which ultimately increased the competition among companies. Success in such a competitive marketplace may indeed rest with how effectively a brand engages with social media. The study seeks to explore the effect of social media marketing efforts on the market penetration of cosmetic firms in the Indian market, focusing on three important aspects which are enhancement of the social image, building relationships, and recognizing the brand. Brand awareness is one of the essential parts of marketing. It is necessary for any business desiring to gain the interest of a customer and retain them. In social media, however, brand awareness encompasses both brand recognition and creating a brand effect through which customers engage or interact with the brand. Social media sites create an exciting environment for a business to be in touch with customers in real time, respond to their queries, and understand their problems. Consumers engaged at this level mean that a brand becomes recognizable to them, and the chances are high that they will opt for it when buying something. Social media gives exposure to up-coming cosmetics, which otherwise may not have wide endorsement or advertisement, to reach a big audience base while reinforcing their brand awareness at low costs.

The relationship development with consumers is another critical concern in an effective social media marketing strategy. Social media allows for two-way communication between brands and customers, unlike traditional marketing plat-forms that are usually one-way. This relationship is particularly helpful in building community and loyalty if cosmetic firms want to have a following that is dedicated. Customized content, direct messaging, and comments bring marketers closer to their audience. This relationship-building feature of social media is very important for the bridging process because early adopters become loyal customers who may influence mainstream consumers in making purchase decisions. The success of social media marketing for the cosmetics business is also based on projecting a positive social image. In a marketplace increasingly concerned about issues of sustainability, variety, and ethical business, a brand's reputation will be considerably linked to how it reflects in the social media arena. Brands that focus more on socially conscious and conduct corporate social responsibility are normally more accepted by mainstream consumers and early adopters. Social media marketing

can be utilized to express projects and views, expand the brand image and attract more consumers.

The purpose of the research is to establish how social media marketing activities are affecting Indian cosmetic brands in effectively bridging the gap that separates early adopters and mainstream consumers is a very critical stage more commonly referred to as "crossing the chasm". The study looks at establishing how social networking sites like Facebook, YouTube, and Instagram will be helpful for cosmetic companies, especially the smaller and more recent ones in the industry, to build closer ties with clients and increase their online visibility in the eyes of the public and raise their presence. The study thus helps one gauge how these digital strategies have enabled cosmetic brands to break into the main-stream market and, subsequently, give them a pathway for sustained growth and competitive advantage in a rapidly shifting Indian cosmetics industry. Despite the growing importance of social media marketing in the cosmetics industry, there remains a significant research gap in understanding how these activities specifically help brands in India bridge the gap between early adopters and mainstream customers. Most existing studies focus on broader digital marketing strategies or consumer behavior in Western markets, with limited attention to the unique dynamics of the Indian market, where cultural nuances, consumer preferences, and digital literacy vary widely. Additionally, while social media's role in increasing brand awareness and customer engagement is well-documented, there is a lack of comprehensive research on how these factors collectively contribute to "crossing the chasm" which is a critical phase for new and small cosmetic brands aiming to achieve mainstream success. This study seeks to fill this gap by providing focused insights into how social media platforms like Instagram, Facebook, and YouTube can be leveraged to navigate the challenges of this transition in the context of India's diverse and rapidly evolving cosmetics market.

The rapid evolution of the Indian cosmetic industry, coupled with the growing influence of social media, presents unique challenges and opportunities for cosmetic brands, particularly in their effort to transition from early adopters to mainstream consumers. This phenomenon can be effectively analyzed through Geoffrey Moore's "Crossing the Chasm" model. Moore [30] posits that there is a significant gap between early adopters and the early majority, which creates a challenge for businesses attempting to scale beyond the innovators and early adopters. Early adopters are typically more experimental, open to new technologies, and willing to engage with novel products. Conversely, the early majority tends to be more risk-averse, requiring solid proof of a product's effectiveness and market stability before committing to a purchase.

In this context, social media marketing (SMM) becomes a critical tool for cosmetic brands in navigating this gap. The visual and interactive nature of platforms like Instagram, Facebook, and YouTube allows brands to create engaging narratives, showcase product effectiveness through influencer endorsements, and build a socially conscious brand image that resonates with both early adopters and the mainstream market [31]. By leveraging SMM, brands can craft a relatable and trustworthy identity, thus easing the transition across Moore's chasm, where the mainstream market demands more than just novelty—they seek proven benefits and emotional connection with the brand.

2 Review of Literature

Social media is a force of change in today's digital times, as much as anything else fundamentally can be, for changing brand consumer interaction across various industries. Indeed, this is well evidenced in the cosmetic industry, where appealing visuals and brand perception are key factors to success [1]. Social media platforms offer opportunities to connect with consumers on personal and interactive levels, defying traditional marketing boundaries. The interactive powers of social media sites allow two-way communication between customers and brands, and thereby building a deeper relationship between them [2]. This has changed the conventional notions of brand loyalty and trust of consumers to rely more on continuing engagement and personal interaction. Brand awareness is considered one of the major bases of marketing communications. It reflects the degree to which consumers can identify or remember a brand. Cosmetics industry is highly competitive in nature. Differentiation and visibility of the products are very important in attracting and retaining customers. The social media played an important role in boosting brand awareness [3]. The social media Platforms such as Instagram, Facebook, and YouTube offer visually engaging and interactive features that enable brands to reach a vast audience and reinforce their presence in the market. Customer Relationship Building focuses on developing and maintaining meaningful interactions between brands and their customers.

2.1 Brand Awareness

Brand awareness is pivotal to marketing success, especially in the cosmetics industry, which completely relies on the appeal and differentiation of one brand from another. Current research has underlined the importance of social media for creating brand awareness and visibility. [5] explored the role of visually oriented Instagram in building brand awareness and pointed out that interactive content with blogger engagement can bring about significant brand awareness. [6] analyzed the impact of coherent social network branding across multiple platforms such as Facebook and Instagram. The results indicated that coherent social net-work branding increases consumers' memory of brands. [7] proved that engaging social network campaigns were among the effective ways to raise brand aware-ness through user involvement in participation and sharing. This is further sup-ported by the work of [8], who showed that the multi-platform approach amplifies brand messages on all different social media channels, and also increases the touchpoints that have been developed for the consumer. [9] reported a positive relationship between social media brand engagement and brand recall, hence active user interactions with the branded content ensured greater brand recall. [10] highlighted that high-quality visual content of social media sites remarkably impacts brand awareness, especially in visually appealing industries like cosmetics. [11] and [12] supported the idea that influencer endorsement importantly helps to enhance brand awareness. Large followings combined with the high credibility of the influencers raise brand visibility and attract more customers.

2.2 Customer Relationship Building

Building strong customer relationships is essential for sustaining loyalty and trust, and social media offers valuable tools for fostering these connections. Effective customer relationship management on social media involves direct communication, personalized content, and proactive engagement. Many studies have highlighted the benefits of real-time engagement on social media, noting that prompt responses to customer inquiries and concerns strengthen relationships [13, 14]. [15] found that personalized interactions and engaging content contribute to higher customer satisfaction and loyalty. [17] emphasized the role of story-telling and brand values in building emotional connections with customers, further enhancing relationship quality. In a study conducted [18], demonstrated that brands that actively engage with their audience through personalized messages and targeted content experience increased customer loyalty and engagement. [19] supported this by showing that creating value-added content that resonates with the audience is crucial for effective customer relationship management.

Relationship Marketing Theory [31] is another key concept that underscores the importance of developing long-term, trust-based relation-ships between brands and customers. Social media has revolutionized how brands and consumers interact, shifting from one-way, top-down communication to two-way, personalized engagements. Cosmetic brands can leverage platforms like Instagram and YouTube to create meaningful, direct interactions with customers, through comments, direct messages, and personalized responses. This form of relationship-building is crucial in bridging the gap between early adopters and mainstream consumers, as it fosters loyalty and encourages early adopters to influence the purchasing decisions of their peers in the mainstream market (Grönroos, 2004).

This shift aligns with Moore's theory, as loyal early adopters can act as brand advocates, helping to "pull" the mainstream market across the chasm by reinforcing the brand's credibility and authenticity through user-generated content and peer recommendations (Moore, 1991).

2.3 Social Identity Theory

Social Identity Theory [34] explains how consumers derive part of their identity from the groups and brands they associate with. In the cosmetics industry, where personal appearance and self-expression are key, consumers are more likely to engage with brands that reflect their values and social identity. With increasing attention to sustainability, inclusivity, and ethical practices in the beauty sector, brands that communicate their commitment to these values on social media are more likely to build a loyal customer base. Brands like The Body Shop and Forest Essentials, for example, have successfully used social media to emphasize their ethical sourcing and sustainability efforts, aligning themselves with socially conscious consumers in India.

This alignment between brand and consumer identity not only helps cosmetic firms bridge the gap between early adopters and the mainstream market but also strengthens the emotional connection between the brand and its customers [35]. By consistently reinforcing a positive social image, brands create deeper loyalty and encourage customers to become advocates who recommend the brand to others, thus broadening the company's reach.

2.4 Social Image Enhancement

Enhancing a brand's social image through social media involves promoting a positive public perception and aligning with consumer values. Social media plat-forms helps brands to showcase their commitment to social responsibility and ethical practices [20, 21]. [22, 23] investigated how social media is used to highlight corporate social responsibility (CSR) initiatives, improving the brand's image and aligning with consumer values and found that brands engaged in social causes and shared these efforts on social media experience enhanced public perception. [24] highlighted the impact of transparency and authenticity in social media communications on a brand's social image and demonstrated that showcasing commitment to sustainability and ethical practices on social media positively impacts brand reputation and consumer trust. [25] explored how positive social media interactions and endorsements contribute to a favorable brand image. [26] emphasized that social media campaigns focused on social good and community involvement effectively enhance a brand's social image, reinforcing the im-portance of aligning brand activities with consumer values.

2.5 Crossing the Chasm

Crossing the chasm involves transitioning from early adopters to mainstream consumers, a challenge that social media can help address by expanding brand reach and building credibility [27]. [28] examined how social media marketing strategies facilitate this transition, including influencer partnerships and targeted advertising and found that integrating innovative content and inclusive messaging appeals to both early adopters and mainstream audiences.[30] highlighted the role of social media analytics in understanding consumer behavior and tailoring strategies to reach mainstream audiences effectively and demonstrated that incorporating user-generated content and customer testimonials into social media strategies helps build credibility and appeal to a broader audience. [31] emphasized the importance of maintaining a strong presence across multiple social media platforms to bridge the gap between early adopters and mainstream consumers (Fig. 1).

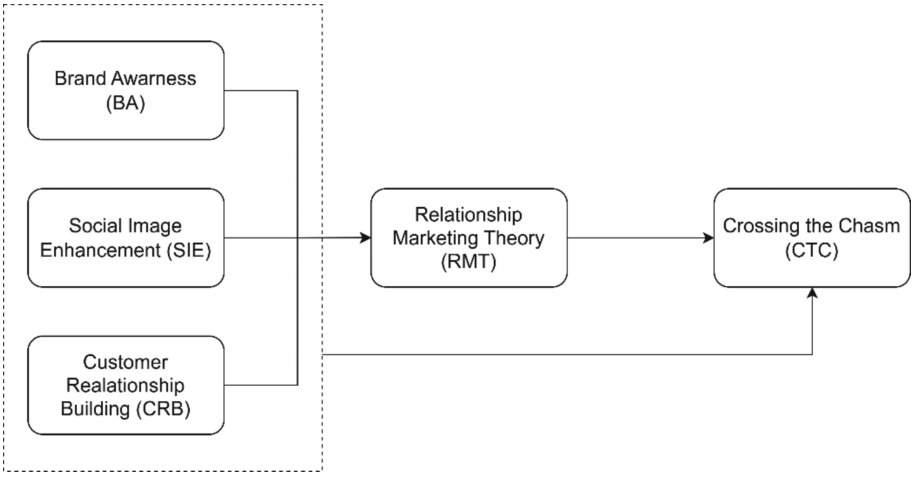


Fig. 1. Theoretical Framework

3 Research Methodology

The survey was done to understand the perception of Social Media Marketing Activities among different target groups within Coimbatore city. The fact that Coimbatore is a metropolitan city with around 4.1 million in the latest Census allowed for a significant spectrum among participants of different age groups and backgrounds in order to ensure a wide range of perspectives. The geo-graphical focus of Coimbatore city was chosen to give better precision in the findings, as this kind of localized study allowed more control and specificity to analyze the SMMA in this very particular context. As a sample size of 424 was identified by means of a sample size calculator, because at least 385 participants were recommended for capturing a representative sample of the population.

This sample size was sufficient to give statistically significant information on the aspects that the study aimed to establish. The respondents were selected using a non-probability sampling method, namely, convenience sampling. Convenience sampling was chosen as a strategy to obtain research participants because of efficiency, an aspect crucial in reaching a sufficient number of participants considered active users of social media-an important criterion for this study.

For the purpose of data collection, a structured questionnaire was used that measured different aspects of SMMA: brand image, customer relationships, and how social media marketing has an influence on customers' behavioral aspects. The questionnaire was pretested through a pilot study comprising of 25 people, to ensure that all the questions were relevant and clear to respondents. From the results of the pilot study, some minor adjustments had been carried out to enhance the readability and completeness of the questionnaire. To ensure that the data collected would be reliable, a reliability check was done through Cronbach's alpha of value 0.80, showing high internal consistency of the measurement scale.

The content and structure of the questionnaire were reviewed by an panel of experts consisting of marketing professionals and academics for validation. The final questionnaire was then distributed online and physically, which gave it greater reach and ensured a pool of diverse participants. The focus was on Coimbatore city because of its large growing urban population, increasingly engaged in social media activities due to which the location is ideal for studying the im-pact of SMMA. For this reason, the population was judged appropriate for the study because respondents from it reflect both traditional and modern consumer behaviors, hence allowing valuable deductions on how SMMA might influence a wide range of consumers.

These represent the results of reliability analysis, as shown in Table 1, to test the internal consistency and validity of the variables under study. It shows that across the variables, there was a high level of reliability because their AVE values were high. AVE values higher than 0.50 mean, more variance is captured by the construct than the error variance, which is a positive indicator of reliability. The 'Customer Relationship' and 'Crossing the Chasm' have high AVEs of 0.698 and 0.741 respectively, thus confirming that these constructs are internally consistent. The 'Positive Social Image' has a higher AVE value of 0.968, showing that it is a very reliable measure. The reliability of the high scores consistently obtained across all the variables reinforces the credibility of the measurement model adopted for this study.

Table 1. Reliability Analysis

Variables	Cronbach's Alpha	McDonald's omega	Average Variance Extract-ed (AVE)
Brand Awareness	0.882	0.884	0.579
Customer Relation	0.907	0.909	0.698
Positive Social image	0.879	0.881	0.968
Relationship Marketing Theory	0.886	0.890	0.581
Crossing the chasm	0.882	0.884	0.741

Table 2. Heterotrait - Monotrait Ratio (HTMT) of correlations

Variables	Brand awareness	Customer Relationship	Positive Social image	Relationship Marketing Theory	Crossing the Chasm
Brand Awareness	1	-	-	-	-
Customer Relation	1.007	1	-	-	-
Positive social image	0.28	0.24	1	-	-
Relationship Marketing Theory	0.13	0.22	0.845	1	-
Crossing the chasm	0.22	0.786	0.658	0.585	1

Table 2 provides the Heterotrait - Monotrait Ratio (HTMT) of correlations, which is used to assess discriminant validity. Discriminant validity is confirmed when the HTMT values are below the threshold of 0.85 or 0.90, depending on the strictness of the criteria. The HTMT values in this study indicate good discriminant validity among most constructs. For instance, the HTMT value between 'Brand Awareness' and 'Customer Relationship' is 1.007, slightly above the threshold, suggesting a need for cautious interpretation. However, most other relationships, such as between 'Brand Awareness' and 'Positive Social Image' (0.28) and between 'Relationship Marketing Theory' and 'Crossing the Chasm' (0.585), are well within acceptable limits. The highest HTMT value is observed between 'Positive Social Image' and 'Relationship Marketing Theory' (0.845), which is on the upper edge of the acceptable range but still indicates adequate discriminant validity.

4 Data Interpretation and Findings

4.1 Descriptive Analysis

In the demographic profile described in Table 3, it was recorded that female respondents outnumbered their male counterparts. For instance, 69.8% were female, while 30.2% were male. This has also been confirmed by the fact that the majority of cosmetic products target this category as their potential customers. The majority in this sample population consists of young adults, where 58.3% of the population is within the age group of 21–25 years and 33.9% of the population within the age range of 18–21 years. Only 7.8% of the respond-ents are aged 25 years and above, which fairly indicates that the study to a large extent captures the stands of younger consumers who are in all likelihood more engaged with digital platforms. Regarding the usage of social media, Instagram floats to the top as the most used social media platform among the respondents at 86.16%, closely followed by YouTube at 84.91%.

Therefore, in order of preference, the platforms are: Instagram, preferred by 58.33% of the interviewed, and YouTube, chosen by 53.85%. Much less popular are the following: Facebook. As clear, this reveals an extraordinary trend toward photograph/video exchanges on Instagram and YouTube, underlining their relevance for cosmetic brands targeting the young audience. Now, regarding the frequency with which social media is used, 69.3% of respondents access social media many times a day and 27.1% a few times a day. Only 2.6% of the respondents go onto their social media once a day, while 0.9% rarely do. The high frequency of this engagement underscores the importance of a strong and continued social media presence that brands would want to attain with this demographic.

Table 3. Demographic Profile of the Respondents

Characteristics	Respondent's Profile	Frequency of Respondents
Gender	Male	30.2%
	Female	69.8%
Age	18 – 21	33.9%
	21- 25	58.3%
	25 years and above	7.8%
Social Media	Instagram	86.16%
	YouTube	84.91%
	Facebook	22.64%
	Twitter	19.50%
Frequency	Multiple times a day	69.3%
	Few times a day	27.1%
	Once in a day	2.6%
	Rarely	0.9%

4.2 Factors that Enhance Social Media Marketing Activities

The analysis of the mean and standard deviation for the key variables reveals a positive perception among respondents regarding the role of social media in the cosmetic industry presented in Table 4. The mean score for Brand Awareness (3.97) indicates a high level of agreement on the importance of social media in enhancing brand recognition, with a standard deviation of 0.565 suggesting consistent views across the sample. Similarly, the Customer Relation variable, with a mean of 3.64 and a standard deviation of 0.490, reflects a generally favorable opinion on the effectiveness of social media in building and maintaining custom-er relationships, though slightly less pronounced than brand awareness. The high mean score for Positive Social Image (3.84) and its low standard deviation (0.485) further underscores the consensus on social media's role in improving a brand's social perception. Relationship Marketing Theory received a mean score of 3.79 with a standard deviation of 0.518, indicating strong agreement on the relevance of relationship marketing principles in the context of social media. Lastly, the variable Crossing the Chasm achieved a mean score of 3.91 with a standard deviation of 0.512, demonstrating robust support for the notion that social media marketing activities are crucial for transitioning from early adopters to mainstream customers. Overall, the data highlights a generally positive and consistent view among respondents on the effectiveness of social media marketing in various aspects of brand management within the cosmetics industry.

Table 4. Mean and standard deviation

Variables	Mean	Standard Deviation
Brand Awareness	3.97	0.565
Customer Relation	3.64	0.490
Positive Social image	3.84	0.485
Relationship Marketing Theory	3.79	0.518
Crossing the chasm	3.91	0.512

The correlation matrix reveals significant positive relationships between the variables examined. Brand Awareness (BA) shows strong positive correlations with Customer Relation (CR) ($r = 0.605$, $p < 0.01$), Positive Social Image (PSI) ($r = 0.728$, $p < 0.01$), Relationship Marketing Theory (RMT) ($r = 0.693$, $p < 0.01$), and Customer Conversion (CC) ($r = 0.716$, $p < 0.01$). These correlations suggest that higher brand awareness is associated with stronger customer relations, a more positive social image, greater alignment with relationship marketing theory, and better customer conversion rates. Customer Relation also exhibits significant positive correlations with Average Positive Social Image ($r = 0.595$, $p < 0.01$), Average Relationship Marketing Theory ($r = 0.629$, $p < 0.01$), and Average Customer Conversion ($r = 0.543$, $p < 0.01$). This indicates that better customer relations are linked to a more favorable social image, a stronger relationship marketing approach, and improved customer conversion outcomes.

Table 5. Correlation Matrix

Variables		Avg BA	Avg CR	Avg PSI	Avg RMT	Avg CC
Avg BA	Pearson's r	-				
	df	-				
	p-value	-				
Avg CR	Pearson's r	0.605	-			
	df	422	-			
	p-value	< 0.01	-			
Avg PSI	Pearson's r	0.728	0.595	-		
	df	422	422	-		
	p-value	< 0.01	< 0.01	-		
Avg RMT	Pearson's r	0.693	0.629	0.668	-	
	df	422	422	422	-	
	p-value	< 0.01	< 0.01	< 0.01	-	
Avg CC	Pearson's r	0.716	0.543	0.784	0.686	-
	df	422	422	422	422	-
	p-value	< 0.01	< 0.01	< 0.01	< 0.01	-

Positive Social Image has high positive correlations with Average Relationship Marketing Theory ($r = 0.668$, $p < 0.01$) and Average Customer Conversion ($r = 0.784$, $p < 0.01$). These results suggest that a positive social image is closely tied to the effectiveness of relationship marketing strategies and higher customer conversion rates. Average Relationship Marketing Theory and Average Customer Conversion are also positively correlated ($r = 0.686$, $p < 0.01$), highlighting that effective relationship marketing contributes to better customer conversion results. The correlations demonstrate that brand awareness, customer relations, positive social image, relationship marketing, and customer conversion are interrelated, with strong, positive associations across all variables. This indicates that improvements in one area are likely to benefit others, reinforcing the importance of a holistic approach to social media marketing.

The path coefficients and P-values in Table 6 provide significant insights into the relationships between various constructs in the context of social media marketing activities (SMMA) in India's cosmetic industry. All the hypothesized relationships were supported with strong path coefficients and statistically significant P-values (<0.01), indicating robust associations. A positive social image significantly impacts relationship marketing theory (path coefficient = 0.30), suggesting that how a brand is perceived socially plays a crucial role in fostering customer relationships. Customer relationships with a 0.42 path coefficient and brand image with a 0.60 path coefficient are also strongly linked to relationship marketing theory, underscoring the importance of these elements in effective relationship marketing strategies.

The positive social image also plays a significant role in crossing the chasm and has a path coefficient of 0.22, which represents the effect it exerts on making a transition from early adopters to the mainstream market. Similarly, customer relationship and brand image, with path coefficients 0.31 and 0.44 respectively, are essential ingredients of such transition and go on to reinforce the necessity for solid customer ties and an appealing brand identity. In this way, the theory of relationship marketing itself has a significant effect on crossing the chasm via the path coefficient of 0.24 and confirms that its notions of relationship marketing become necessary in order to help the brands cross over from niche markets to wider acceptance. Generally, all these findings pinpoint how the social image, customer relationship, brand image, and marketing strategies all interlink within the successful extension of cosmetic brands by social media marketing (Fig. 2).

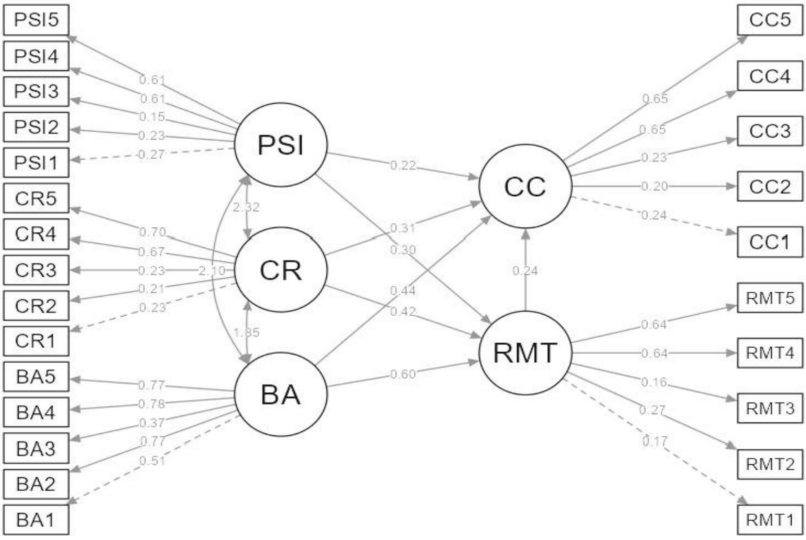


Fig. 2. Path Diagram

Table 6. .

Path	Path Co-efficient	P Value	Hypothesis Decision
Positive social image - Relationship marketing theory	0.30	< 0.01	Accepted
Customer Relationship - Relationship marketing theory	0.42	< 0.01	Accepted
Brand image - Relationship marketing theory	0.60	< 0.01	Accepted

(*continued*)

Table 6. (*continued*)

Path	Path Co-efficient	P Value	Hypothesis Decision
Positive social image - Crossing the chasm	0.22	< 0.01	Accepted
Customer Relationship - Crossing the chasm	0.31	< 0.01	Accepted
Brand image - Crossing the chasm	0.44	< 0.01	Accepted
Relationship marketing theory - Crossing the chasm	0.24	< 0.01	Accepted

5 Conclusion

The study clearly highlights the social media marketing approach SMMA and how it affects the cosmetic industry marketing in India with special focus on helping industry players to bridge the gap between early adopters and the rest of the population. Perceptions of the customers of Coimbatore city have also contributed towards insights in this study were customers can help enhance negative brand image, brand loyalty and relationships in the future use. It is evident from the findings that social media marketing activities SMMA play a positive role in strengthening the brand and building customer dedication which is crucial in marketing in a competitive environment.

The findings underscore the significance of the instruments required to enhance the images functioning inside reality to improve the probability of market entry. These insights are particularly beneficial for emerging cosmetic firms seeking to mitigate the challenges of expanding their market presence. The study enhances the understanding of how social media marketing influences key aspects of customer behavior, including market dispersion. When undertaking further research, it is essential to consider exploring additional aspects of this task, since it has the possibility for completion. An illustration of this is the transformation of nations into extensive markets for marketing via SMMA. Assessing the success of SMMA across several areas or sectors might facilitate further study, enhancing the practice's effectiveness in diverse contexts. This research can be augmented with other inquiries. This study's findings illustrate the importance of social media marketing as a technique for enhancing enterprises and cultivating markets in contemporary society.

5.1 Research Implication

The research has noteworthy contributions to the marketing profession and academic circles. For practitioners, especially those in the cosmetic industry, the study underscores the importance of social media marketing activities (SMMA) in promoting brand awareness, developing customer trust/relationships and serving a favorable image. Such factors help in closing the gap between those who buy first and the rest of the population and therefore are useful for growth in a market. Brands are able to use these to come up with specific social media campaigns that enhance the brand as well as the audience's

engagement with the brand. The focus on local consumers' attitudes likewise provides useful information on how to go about U.S. market entry more precisely.

For authors, this research adds to the corpus of knowledge in digital marketing as it provides an insight into the effects of SMMA on consumer behavior and market penetration. The article contains a thorough methodological framework with a clear focus on reliability and validity, which may serve as an example for other studies in the area. Thus, other researchers may be interested in these hypotheses and delve into the use of SMMA in other fields or countries perhaps revealing other patterns and approaches. Also, such knowledge and insights from this research can contribute to the theory of relationship marketing and consumer behavior thereby expanding scholarly discussions regarding digital marketing approaches.

5.2 Limitation and Future Direction

The research describes the role of social media marketing activities in the cosmetic industry, but most of it was based on the city of Coimbatore and therefore may lack the diversity of consumer behavior in any other region. It is potential that the result would be affected by its local market characteristics which has limited generalization for a broad con-text. Besides, convenience sampling, although easy to operate, may introduce bias and influence the sample's representativeness. It is possible to extend this research to more cities or regions to better understand the effects of SMMA in different markets. Future research might also use probabilistic sampling to enhance the representative nature of their findings. This again means that further research in other industries or sectors could show more specifically how different SMMA applications come into play. Long-term analysis of the influence that social media strategies have on brand loyalty or consumer behavior would also lead to an increased understanding of sustained effects. In these ways, too, such research directions should be expected to yield a more sophisticated and generalizable body of knowledge in digital marketing practices.

References

1. Ho, J.S.: Effects of social media on cosmetic brands. J. Market. Adv. Pract. **1**(2), 1–10 (2019)
2. Maitri, W.S., Suherlan, S., Dwi, R., Prakosos, Y., Subagja, A.D.: Recent trends in social media marketing strategy. J. Minfo Polgan **12**(2), 842–850 (2023)
3. Chinomona, R.: The influence of brand awareness, brand association and product quality on brand loyalty and repurchase intention: a case of male consumers for cosmetic brands in South Africa (2017)
4. Marc, W., Rasul, T.: Customer engagement and social media: revisiting the past to inform the future. J. Bus. Res. **148**, 325–342 (2022)
5. Tasia, A., Indra, R.: Exploring Instagram's visual communication effects of Indonesia's premium fashion brands. **100**, 1 (2022)
6. Rijitha, R.: The impact of social media marketing on consumer purchase (2023)
7. B. Awareness and B. Image, IMAR Indonesian management and accounting research the role of social media marketing in increasing, **20**, 02 (2021)

8. Siregar, N., Nursyamsi, S.E., Angellia, F., Jashinta, M., Hamboer, E.: The role of social media in increasing customer interaction and brand loyalty. Jurnal Minfo Polgan **12**(2), 1865–1873 (2023)
9. Cheung, M.L., Pires, G., Rosenberger, P.J.: The influence of perceived social media marketing elements on consumer–brand engagement and brand knowledge. Asia Pac. J. Mark. Logist. **32**(3), 695–720 (2020)
10. Khoshtaria, Todua, N.: The impact of social media influencers on brand awareness, image and trust in their sponsored content : an empirical study from Georgian social media users. Int. J. Marketing, Commun. New Media **10**, 18 (2022)
11. Dwivedi, Y.K., et al.: The future of digital and social media marketing research: perspectives and research propositions. Int. J. Inf. Manage. **59**, 2020, 102168 (2021)
12. Liu, X., Zheng, X.: The persuasive power of social media influencers in brand credibility and purchase intention. Human. Soc. Sci. Commun. **11**(1), 1–12 (2024)
13. Rane, N.L., Achari, A., Choudhary, S.P.: Enhancing customer loyalty through quality of service: effective strategies to improve customer satisfaction, experience, relationship, and engagement. Int. Res. J. Modernization Eng. Technol. Sci. **5**(5), 427–452 (2023)
14. Shaw, S.D., Carpenter, S.M., Murphy, S.A.: Engagement in digital interventions. Am. Psychol. **77**(7), 836 (2022)
15. Cardona, M.A.D., Ishmael, K.: Artificial intelligence and the future of teaching and learning (2023)
16. Teraiya, V.S., Jayswal, R.: The impact of storytelling advertising on consumer advertising experiences and brand equity. GLIMS J. Manage. Rev. Trans. **2**(1), 1–23 (2023)
17. Chandra, S., Verma, S., Donthu, N., Lim, W.M.: Personalization in personalized marketing: trends and ways forward. Psychol. Mark. **39**(8), 1529–1562 (2022)
18. Nasır, S.: Customer relationship management as a customer-centric business strategy. In: Advertising and Branding: Concepts, Methodologies, Tools, and Applications, pp. 649–685. IGI Global (2017)
19. Jeswani, R.: The role and importance of social media marketing in brand building. **7**(4), 1–9 (2023)
20. Wang, S., Liao, Y., Wu, W., Bao, K., Le, H.: The role of corporate social responsibility perceptions in brand equity, brand credibility, brand reputation, and purchase intentions. Sustainability **13**(21), 11975 (2021)
21. Liu, Y., et al.: Psychology in action: social media communication, CSR, and consumer behavior management in banking. PLoS ONE **18**(8), e0289281 (2023)
22. Simone, L., Ballerini, J., Santoro, G., Dabi, M.: Consumer engagement through corporate social responsibility communication on social media: evidence from Facebook and Instagram Bank Accounts. J. Bus. Res. **172**, 114433 (2024)
23. Battocchio, A.F.: Effects of transparent brand communication on perceived brand authenticity and consumer responses. J. Product Brand Manage. **30**(8), 1176–1193 (2021)
24. Jamil, K., et al.: Role of social media marketing activities in influencing customer intentions: a perspective of a new emerging era. Front. Psychol. **12**, 808525 (2022)
25. Barger, V.A., Peltier, J.W.: Social media and consumer engagement: a review and research agenda social media and consumer engagement: a review and research agenda. J. Res. Interact. Mark. **10**(4), 268–287 (2016)
26. Majka, M., Poland, N.T.: Crossing the Chasm: navigating the leap from early adopters to mainstream success crossing the chasm: navigating the leap from early adopters to mainstream success (2024)
27. Salhab, H.: Social media marketing — Rise of social media influencer marketing and Instagram. J. Infrastruct., Policy Dev. **8**(8), 3783 (2024)
28. Chan, F.: A study of social media influencers and impact on consumer buying behaviour in the United Kingdom. Int. J. Bus. Manage. Stud. **3**(07), 2694–1449 (2022)

29. Rosário, A.T., Europeia, U.: Marketing strategies on social media platforms. Int. J. E-Bus. Res. (IJEBR) **19**(1), 1–25 (2023)
30. Moore, G.A.: Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers. HarperBusiness (1991)
31. Kaplan, A.M., Haenlein, M.: Users of the world, unite! The challenges and opportunities of social media. Bus. Horiz. **53**(1), 59–68 (2010). https://doi.org/10.1016/j.bushor.2009.09.003
32. Morgan, R.M., Hunt, S.D.: The commitment-trust theory of relationship marketing. J. Mark. **58**(3), 20–38 (1994). https://doi.org/10.1177/002224299405800302
33. Grönroos, C.: The relationship marketing process: communication, interaction, dialogue, value. J. Bus. Ind. Marketing **19**(2), 99–113 (2004). https://doi.org/10.1108/08858620410523981
34. Tajfel, H., Turner, J.C.: The social identity theory of intergroup behavior. In: Worchel, S., Austin, W.G., (Eds.), Psychology of Intergroup Relations (2nd ed., pp. 7–24). Nelson-Hall (1986)
35. Choi, S.M., Rifon, N.J.: It is a match: The impact of congruence between celebrity image and consumer ideal self on endorsement effectiveness. Psychol. Mark. **29**(9), 639–650 (2012). https://doi.org/10.1002/mar.20550

Ubiquitous Computing of Low-Cost Pollutant Sensors for Monitoring Pollutant Gases in the Construction Sector

Rebakah Geddam(✉), Mohd Zuhair, Kush Shah, Karm Vyas, and Naitik Kanani

CSE Department, School of Technology, Nirma University, Ahmedabad, India
{rebakah.geddam,mohd.zuhair,19bec120,22BCE137,
22BCE145}@nirmauni.ac.in

Abstract. The growing urbanization in India has been closely associated with the expansion of the cement industry. Acknowledging the significant impact of air pollution from the cement industry on the environment, human health, and the economies of affected nations is imperative. To safeguard individuals from health issues arising from air pollution, it is crucial to accurately and promptly anticipate the ambient air quality index. The pollution control board conducts an annual audit to ensure the implementation of Sustainable Development Goals (SDG) policies aimed at climate mitigation. A low-cost sensor employing a cloud-edge architecture was installed in the cement industry to monitor Ambient Air Quality for ubiquitous computing. The innovative approach in this study involves monitoring air quality by detecting anomalies during a calibration drive. Comprehensive emissions monitoring is achieved across various operational scenarios by combining data from multiple sensors, including pollution sensors, PPM sensor arrays, and environmental sensors. The results demonstrate a substantial enhancement in the accuracy of air quality monitoring and the early detection of potential health risks. The article elucidates anomaly detection algorithms, namely Isolation Forest (IFor), Local Outlier Factor (LOF), and DBSCAN, compared between the industry-set sensor *sn* and the environment lab-tested referenced sensor *sv*. After removing anomalous readings, RMSE scores of 22.75 and 17.7 were recorded. The article emphasizes adaptable solutions for avoiding unhealthy air, benefiting the environment, and improving public health.

Keywords: Sensor Drift · Anomaly Detection algorithm · Ambient Air Quality (AAQ) · Environmental Sensor · Cement Industry · Public Health

1 Introduction

The economic growth in Gujarat has observed an exodus of rural areas to urbanization. There is a demand for industrialization, infrastructure, improved lifestyle, resources, electricity, and raw materials. The real estate sector took an elevated turn to satisfy the urbanization needs for an improved lifestyle, better infrastructure, and transportation. The construction of infrastructure and transportation and the supply of raw materials

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 425–434, 2025.
https://doi.org/10.1007/978-3-031-86305-9_29

come from the cement industry. Gujarat witnessed a 39.08% rise in cement aggregate production [1]. Preserving a sustainable environment is necessary as industrial and infrastructural growth occurs within the region. The cement industry is one of the significant contributors to greenhouse gases and global warming. Cement is manufactured using non-metallic minerals such as limestone and coal. The third-largest industrial contributor of carbon pollution is the cement industry, emitting more than 500,000 tons per year of $CO2$ (carbon dioxide), $SO2$ (sulfur dioxide), $NO2$ (Nitrogen Dioxide), and CO (Carbon Monoxide) [2]. Along with the gases, air pollutants like particulate matter ($PM2.5$, $PM10$) are released during cement manufacturing. The pollutants consist of microscopic particles which are inhaled directly into the lungs. The toxic airborne pollutants have adverse effects on the health of the people working in the industry and the nearby residential localities. The concentrations of these pollutants in the air have the effect of headaches, anxiety, and cardiovascular disease for the factory workers and occupants. In the industrial area, the Cement industry is directly associated with the mortality rate concerning coronary heart disease and lung cancer [3]. World Health Organization has suggested a threshold value for $PM2.5$ concentration in air, which is 1.9 times above the guidelines [4] in Gujarat.

1.1 Standards of AAQ Index

The World Health Organization has introduced the Ambient Air Quality (AAQ) parameter to measure the severity of inhalation pollution in the air. India's AAQ is categorized into five levels: 0–50, 51–100, 101–200, 201–300, 301–400, and 401–500, indicating the impact on human health. The construction sector has to maintain the AAQ value above 100, emphasizing satisfactory and moderate. The annual audit for the AAQ index once every three months is conducted by the GPCB [5]. GPCB (Gujarat Pollution Control Board) audits following SDG (sustainable Development Goals) for the construction industry combating climate mitigation. A critical analysis was conducted by the government body to provide a sustainable environment for the workers. Factory authorities must satisfy the SDG assessment, failing which consequences will be faced [6–9].

1.2 Contributions

This research aims to detect anomaly values in sensor drifting in the cement manufacturing industry using an environment-referenced sensor. Four aspects are explored: a cloud-edge architecture for ambient air quality calculation, an anomaly detection algorithm for sensor drift, a classification algorithm for anomalous and nominal points, adaptable recommendations for future detection, and precautionary measures to protect the environment and public health.

2 Ubiquitous Computing Architecture for Healthy Environment

Particles in the air with a diameter of less than 10 mm (micrometers) can freely enter deep lungs and cause problems with health. Moreover, an increase in relative humidity and dust floating in the air can significantly affect the levels of other dangerous pollutants

in the air. Following the National Building Codes (NBC), which were published in 2016, the permissible limits for primary pollutants (suspended particulate matter) are 10 g/m^3 (annual mean) and 20 g/m^3, respectively.

The authors in [10] analyzed the impact of wind on air quality using data from meteorological stations and monitoring stations for $NO2$ and $PM2.5$ concentrations. The findings suggest that population density substantially affects AQI in cities with better air quality. When comparing the impact of using Fuel Standard V petrol and diesel for the subway, authors found that Fuel Standard V Diesel reduced four pollutants and CO compared to petrol. The stacking ensemble frequently outperforms other machine learning methods for the values of R2 and RMSE, according to studies that assessed how well they performed in predicting air quality. At the same time, AdaBoost provides the best results for MAE [11]. Sensor drifting happens when a low-cost sensor gives some anomaly values [5]. The Detection of some anomalies data present in the process of calculating sustainable air quality was suggested in the study [12]. They have extensively worked on anomaly detection on smart city systems as wireless sensor networks. The studies have compared anomaly detection algorithms for real-time-based classification and prediction. The study analyzed and divided the levels of various air pollution factors to detect the outlier or anomalous locations based on the AAQ [13]. To see anomalous locations from pollution databases, support vector machines, neuro-fuzzy methods, and machine learning architectures of neural networks have all been implemented. The MATLAB tool was applied to obtain the simulation results. However, the scope of security provision through security protocols has yet to be explored. Anomaly detection is when the sensor is calibrated or doesn't function properly, giving an abnormal value. These values help in altering the AAQ index [14].

A research study by the authors [15] examined the adaptive capacity of 12 indicators for air pollution and found that air pollution significantly affects health at an individual level. The sensor location also plays a crucial role in accurately measuring AQI, and using a Dual sensor combination can improve effective measuring. The study also investigated the prediction of AQI for different levels using fuzzy logic set theory. It concluded that an intelligent environmental decision support system (IEDSS) for air-quality prediction has proven good results [16].

AAQ was calculated on different parameters, and Table 1 based on the extensive analysis, some observations have pointed out the drawbacks, to which solutions are proposed in the present research.

1. Many authors have used univariate data predominantly $P\ M2.5$ or $P\ M10$. Some authors [5, 15] calculated the air quality value and classified it as per the standard labels. However, the coefficients between the other air pollutants were not analyzed effectively.
2. The observations in the cement industry reflected a healthy air quality index when the real scenario was contradicting, so the proposed model to countercheck the contradictions in sensor readings, anomaly algorithms were implemented, and Table 1 emphasizes authors as in [16] that considered the anomaly detection algorithms.
3. The present research concentrates on using multivariate data of all the air pollutants and implementing anomaly detection concerning the standard sensor obtained from the environment chamber.

Table 1. Analysis of prior work done on AAQ

	Features	[17]	[18]	[15]	[19]	[20]	[21]	[7]	[14]	[3]	[17]	[22]	[12]	[23]	[21]	[24]	[16]	[25]	[5]	[2]	Ubiquitous	Model
Major Pollutants	O_3	✓			✓											✓	✓					✓
	NO_2	✓			✓	✓					✓	✓			✓	✓	✓					✓
	SO_2	✓			✓						✓	✓			✓	✓	✓					✓
	Lead (*P b*)														✓		✓					✓
	PM_{10} and $PM_{2.5}$	✓	✓		✓	✓		✓	✓		✓	✓	✓		✓	✓	✓	✓				✓
	CO	✓				✓					✓		✓		✓	✓	✓					✓
Environment	Indoor	✓		✓	✓	✓		✓	✓	✓			✓	✓				✓	✓	✓		
	Outdoor	✓	✓						✓		✓	✓			✓			✓		✓		✓
ML Algorithms	Regression Algorithms			✓	✓	✓		✓		✓	✓		✓				✓					
	Classification Algorithms			✓	✓	✓				✓			✓				✓					✓
	Time Series Algorithms			✓													✓					
	Clustering Algorithms																✓					
	Anomaly Detection Algorithms																✓					✓

2.1 Anomaly Detection Model for Pollution Sensors

AAQ kit was deployed in the cement industry in Gujarat, India, to monitor the AAQ index. A cloud-edge architecture having a sensor network *Sn* has been installed at the cement manufacturing plant to measure the AAQ of the air around the plant. The sensors are installed in the computational network to communicate with each other, and edge computing performs necessary actions on the inputs acquired. The proposed model is a 3-tier cloud-edge computing model, and each tier is assigned to a pre-defined task as shown in Fig. 1. The layers of the architecture have been defined through the job performed and the components involved in that particular layer. The model defines 3 layers:' Data Acquisition Layer', 'Data pre-processing Layer', and 'Actuation Layer'. The sensor networks *Sn* and *Sv* are present in the data acquisition layer connected through a sensor in the cement industry. The data is acquired from the sensors and interfaced to the cloud. In the cloud, the processing of the data is performed through which the behavior of the data is predicted. The predicted behavior determines the actuator's output at the actuation layer.

Detecting an anomaly requires deploying anomaly detection algorithms. An applied supervised anomaly detection algorithms for intrusion detection and anomaly detection in the data set [12]. However, the scope of the papers is either limited to security provision through machine learning or performing anomaly detection on the data set for classifying the data. The data is extracted from the sensors, stored at the cloud platform through required communication protocols, and is either visualized or processed to give it as input to the sensors. Work on a cloud network and its communication protocols for reading the sensor data and monitoring the AAQ or AQI for the environment [14]. Communication protocols like LoRaWAN or LoRa have been used for long-range communication between the cloud and sensor networks. LoRa stands for Long Range Radio, a proprietary technology developed for long-range communication for data containing very little information or low-bit-rate communication.

2.2 Data Acquisition Layer

This layer comprises data collected from the sensor networks *Sn* and *Sv*. The sensor network α comprises the various sensors mentioned in Eq. 3–6, where each sensor measures the parameter or the gas mentioned in its title from its surrounding environment.

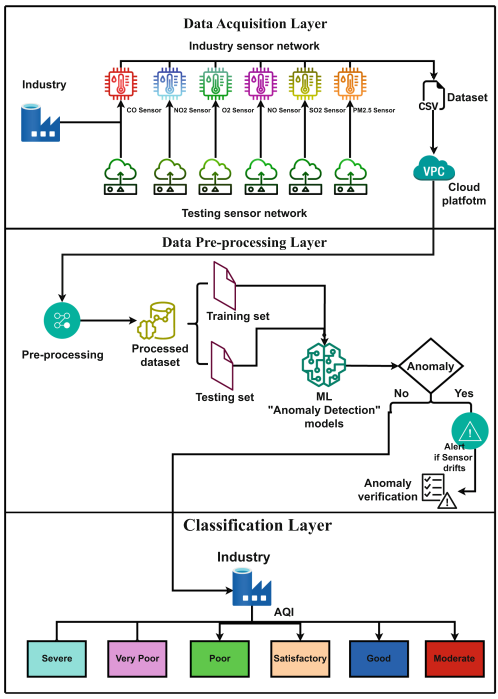


Fig. 1. Ubiquitous computational model for AAQ Monitoring

The data obtained in both the sensor networks, i.e., the industry sensor network and the testing sensor network, is compared and then stored in the cloud for security, verification, and processing purposes. As the sensors can generate output based on the values sensed and communicated amongst the network, the MQTT protocol has been used to establish the communication network amongst the sensors. There might be a possibility of sensor drifting, which is maintaining the AAQ level, as the sensor network *Sn* installed in the industry might be passing on incorrect information. The testing sensor network detects such behavior in the acquisition phase to avoid such scenarios. Once the information is verified, the data is collected in a '.csv' file format and passed on to the cloud regularly. As there is a requirement for anomaly detection, regression analysis, and classification, a secure platform, a virtual private cloud, is chosen so that no data tampering occurs. The data enters the cloud, and the tasks are continued and executed at the intelligence layer for the prediction of the behavior of the sensor network.

The values generated from the sensor network *Sv* have been given different designation symbols.

$$a_1 = NO_2 \tag{1}$$

$$a_2 = O_3 \tag{2}$$

$$a_3 = NO \tag{3}$$

$$a_4 = SO_2 \tag{4}$$

$$a_5 = PM_{2.5} \tag{5}$$

$$a_6 = PM_{10} \tag{6}$$

It can be formulated here that:

$$\forall a \in \{S_V, S_N\} \tag{7}$$

$$\forall a \ni \{a_1, a_2, a_3, a_4\} \tag{8}$$

Each sensor's sub-index range is calculated, and then the non-zero values are added to calculate the AAQ. Further, it is classified based on the range mentioned in the introduction to find the impact of the AAQ. To verify the readings of the sensors, the data is stored in the virtual private cloud to verify whether the readings obtained from *Sn* are malicious. The values obtained from *Sn* are verified with the values obtained by *Sv*. If the readings are not manipulated, the values match and prove that the network *Sn* is not malicious. The function of the system can be formulated as F such that:

$$F = \sum_{i=1}^{n} (RemoveAnomaly(max(verification))) \tag{9}$$

The function 9 maximizes the sensor's verification and removes anomalies. The proposed ubiquitous computing model explains the complete flow of the sensor network verification. Further, the security and verification of the data are also proposed in the system architecture, regression algorithms are applied, and security is provided through the cloud. Through anomaly detection, the malicious data is removed, and through regression and classification of the data, the AAQ of the environment is obtained. Through the output obtained through the classification, the necessary actions are taken. Further, the data is verified and secured in the cloud to check the validity of the sensor.

2.3 Actuation Layer

This layer is defined as the responsibility of performing necessary actions through the actuators (humidifiers, lower carbon fuel) to control the environment of the cement manufacturing plant. The input is obtained from the second layer in the form of predicted and classified values of AAQ. Based on the AAQ condition, the necessary outcomes are taken. If the AAQ values fall above 200, then the construction authorities have recommended lowering the emission of greenhouse gases. An alert is sent to the concerned authorities when the sensor drifts, along with the real-time duration values of the sensor.

3 Results

This section covers the interpolation technique on the data set acquired from the intelligent layer. The null values, incomplete and improper values are filled using the interpolation. The feature extraction method is discussed in the sections below to find the correlation between pollution and ambient air quality and the eigenvalues.

3.1 Correlation Between the Pollutant Variables

The correlation graph in Fig. 2 stated on the X-axis has the names of pollutants, and labels of the Y-axis have the calculated and under consideration parameter, AAQ figured. This graph shows the relationship between AAQ and various other continuous variables. The taller the bar, the more correlation that parameter has with AAQ; lower bars indicate lesser impact. The bars below the x-axis show a negative correlation, as explained in the above section. In this Fig. 2 we can see $O3$, $PM2.5$, and $PM10$ have the top 3 highest bars.

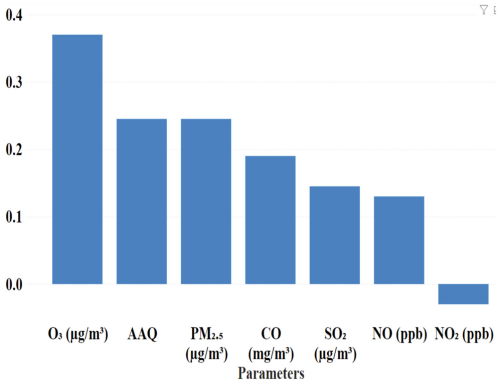
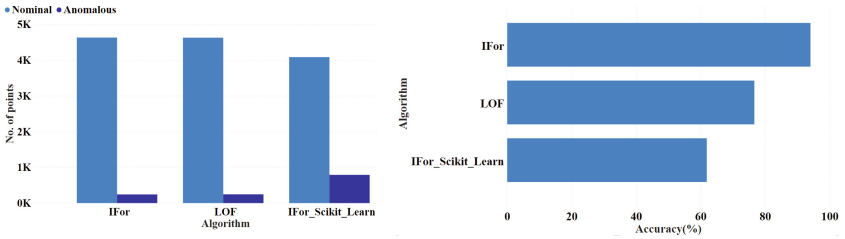


Fig. 2. AAQ correlation with other features.

3.2 Classification of the Anomalous and Nominal Points

Sensor drift is monitored using anomaly detection algorithms to classify the data into anomalous and nominal points. In the intelligence layer, anomaly detection is performed using algorithms isolation forest (IFor), local outlier factor (LOF), and DBSCAN. Figure 3a represents the strange and nominal data points each algorithm finds. The x-axis defines the model applied, while the y-axis depicts the number of minor and uncommon issues, where the red-colored bar is an anomaly, and the blue-colored bar represents insignificant points. The algorithms have detected 241, 244, and 788 abnormal data points as monsters moving left to right in the graph. As shown in Fig. 3, the accuracy achieved by each anomaly detection model. The x-axis represents the model's accuracy, while the y-axis represents the model applied. The accuracy was based on the good, satisfactory, moderate, and poor AAQ points it retained. All the models have successfully removed the poor and very poor data points of AAQ. The accuracy obtained for the models is 61.89%, 76.65%, and 94.035%, moving top to bottom in the graph, respectively. The algorithm of IFor has a flow and base similar to the random forest algorithm, where the points are segregated in the form of a tree until the data point is found. Similarly, for the local outlier factor, the base is similar to the algorithm of k-nearest neighbors, where the model's behavior is predicted based on its neighbors. Represent the anomalies found at the time instances for the calculated AAQ. The x-axis is the time frame of that

particular day, the primary y-axis is the calculated AAQ, and the secondary y-axis is the behavior of the data point, whether it is anomalous or not.



(a) Time Frame 1 (b) Time Frame 2

Fig. 3. (a), (b) Time frame Analysis with anomaly points.

3.3 Regression-Based Results

The regression analysis is performed to predict the value of the calculated AAQ. One of the performance parameters for regression is the Root Mean Square Error or RMSE score, which is similar to the term accuracy for the classification models. RMSE can be formulated as:

$$RMSE = \sqrt{\sum_{i=1}^{N} \frac{(n_i - n_i')^2}{n}} \tag{10}$$

Where 'N' is the number of data points in the dataset, and 'n' is the particular data point selected for processing. The lower the RMSE score, the better the performance of the model. The algorithms iterated are light gradient boosting machine (LGBM), random forest regressor (RFR), gradient boosting regressor (GBR), and decision tree regressor (DTR), whose root mean square value before and post anomaly detection as in Table 2.

The best model for the regression can be stated as the random forest regressor (RFR) as the before and the post anomaly detection. The RMSE score is the lowest among the models applied.

Table 2. Comparison of RMSE Pre and Post Anomaly Detection

Algorithms Sensor drift	No Sensor Drift	
LGBM	22.97	18.03
RF	22.75	17.7
GBR	24.07	19.5
DT	32.06	24.8

4 Conclusion

When a calibration drive is conducted in request with the cement industry management, the sensor deployed in the cement construction sector has witnessed sensor drift readings. Measuring the Ambient Air Quality (AAQ) revealed inappropriate air index readings compared to environmental referenced sensor observations. It was systematically assessed for its ability to detect ambient $PM10$ and $PM2.5$ distributions, employing multivariate interpolation and anomaly detection algorithms. The suggested system oversees the AAQ of the cement plant's air quality, scrutinizing whether the values are accurately represented. A sensor network verification and calibration sensor have been deployed to counteract anomalous readings. Regular calibration, appropriate upkeep, and validation against established criteria can reduce low-cost sensor drifting issues. A classification-based prediction for AAQ values, sensed within the range of 0–400, has been graphed. The outcomes underscore that the model identifies anomalies from the low-cost sensor and pinpoints abnormal data points. These outliers are subsequently identified from the data set, leading to the re-calibration of AAQ. Additionally, the accuracy rate and Root Mean Square Error (RMSE) score for the anomaly detection algorithms and regressors were predicted, revealing that the "IFor" and Random Forest algorithms performed the most reliably. The study also considers the implications of sensor-based anomaly identification in the direction of health protection, highlighting the possibility of proactive measures to reduce risks to public health and the environment. Future work will encompass the integration of security protocols and the establishment of real-life instances for comprehensive implementation. Furthermore, techniques for sensor calibration will be examined to draw comparisons between industrial-grade and testing sensor networks.

References

1. Reddy, Y.K.: Production efficiency of Indian cement industry. Int. J. Manag. (IJM) **11**(11), 2731–2737 (2020)
2. Abulude, F.O., Akinnusotu, A., Bello, L., Feyisetan, A.O.: Assessment of AQI, pm10, pm2. 5, no2, o3: the case of OWO, Nigeria. ASEAN J. Sci. Eng. **4**(1), 15–24 (2024)
3. Jariwala, D., Christian, R.: Development of indoor work environmental air quality index for a dyeing and printing industry. J. Air Pollut. Health (2023)
4. Goshua, A., Akdis, C.A., Nadeau, K.C.: World health organization global air quality guideline recommendations: executive summary. Allergy **77**(7), 1955–1960 (2022)
5. Fung, P.L., et al.: Improving the current air quality index with new particulate indicators using a robust statistical approach. Sci. Total Environ. **844**, 157099 (2022)
6. Singh, N., Gupta, K., Kapur, B.: Greenwashed word of mouth (gwwom): a vibrant influence on customer green behaviour. J. Glob. Responsib. **13**(4), 472–487 (2022)
7. Venkata Sudhakar, C., Reddy, U.: Impacts of cement industry air pollutants on the environment and satellite data applications for air quality monitoring and management. Environ. Monit. Assess. **195**(840), 840 (2023)
8. Sung, W.-T., Hsiao, S.-J.: Building an indoor air quality monitoring system based on the architecture of the internet of things. EURASIP J. Wirel. Commun. Netw. **2021**, 1–41 (2021)
9. Niu, Y., et al.: Meta analysis of heavy metal pollution and sources in surface sediments of lake Taihu, China. Sci. Total Environ. **700**, 134509 (2020)

10. Seo, J., Park, D.-S.R., Kim, J.Y., Youn, D., Lim, Y.B., Kim, Y.: Effects of meteorology and emissions on urban air quality: a quantitative statistical approach to long-term records (1999–2016) in Seoul, South Korea. Atmos. Chem. Phys. **18**(21), 16121–16137 (2018)
11. Tariq, H., Abdaoui, A., Touati, F., Al Hitmi, M.A., Crescini, D., Mnaouer, A.B.: Real-time gradient-aware indigenous AQI estimation IoT platform. Adv. Sci. Technol. Eng. Syst. J. **5**, 1666–1673 (2020)
12. Jain, R., Shah, H.: An anomaly detection in smart cities modeled as wireless sensor network. In: 2016 International Conference on Signal and Information Processing (IConSIP), pp. 1–5. IEEE (2016)
13. Nguyen, D.-D.: Cloud-based drone management system in smart cities. In: Development and Future of Internet of Drones (IoD): Insights, Trends and Road Ahead, pp. 211–230 (2021)
14. Wang, Y., Huang, L., Huang, C., Hu, J., Wang, M.: High-resolution modeling for criteria air pollutants and the associated air quality index in a metropolitan city. Environ. Int. **172**, 107752 (2023)
15. Bildirici, M.E.: The relationship between cement production, mortality rate, air quality, and economic growth for China, India, Brazil, Turkey, and the USA: MSCBVAR and MSCBGC analysis. Environ. Sci. Pollut. Res. **27**(2), 2248–2263 (2020)
16. Dutta, D., Pal, S.K.: Z-number-based AQI in rough set theoretic framework for interpretation of air quality for different thresholds of pm2. 5 and pm10. Environ. Monit. Assess. **194**(9), 653 (2022)
17. Aladag, E.: The influence of meteorological factors on air quality in the province of van, turkey. Water Air Soil Pollut. **234**(4), 259 (2023)
18. Asghar, K., et al.: Assessment of particulate matter (pm) in ambient air of different settings and its associated health risk in Haripur City, Pakistan. Braz. J. Biol. **84**, e256190 (2022)
19. Chen, L.-J., Ho, Y.-H., Hsieh, H.-H., Huang, S.-T., Lee, H.-C., Mahajan, S.: ADF: an anomaly detection framework for large-scale pm2. 5 sensing systems. IEEE Internet Things J. **5**(2), 559–570 (2017)
20. Xu, T., Yan, H., Bai, Y.: Air pollutant analysis and AQI prediction based on GRA and improved SOA-SVR by considering covid-19. Atmosphere **12**(3), 336 (2021)
21. Sayad, T., Moursy, F.I., El-Tantawi, A.M., Saad, M., Morsy, M.: Assessment the impact of different fuels used in cement industry on pollutant emissions and ambient air quality: a case study in Egypt. J. Environ. Health Sci. Eng. **21**(1), 107–121 (2023)
22. Akanksha, A., Maurya, N., Jain, M., Arya, S.: Prediction and analysis of air pollution using machine learning algorithms. In: 2023 3rd International Conference on Intelligent Technologies (CONIT), pp. 1–6. IEEE (2023)
23. Liebert, A., Weber, W., Reif, S., Zimmering, B., Niggemann, O.: Anomaly detection with autoencoders as a tool for detecting sensor malfunctions. In: 2022 IEEE 5th International Conference on Industrial Cyber-Physical Systems (ICPS), pp. 01–08. IEEE (2022)
24. Rani, A., Prakash, V., Darbari, M.: Analysis of ambient air quality by exploring lorawan based connectivity in fog-based infrastructure. In: 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), pp. 145–151. IEEE (2022)
25. Han, Z., Ma, H., Shi, G., He, L., Wei, L., Shi, Q.: A review of groundwater contamination near municipal solid waste landfill sites in china. Sci. Total Environ. **569**, 1255–1264 (2016)

Envisioning the AI-Driven Future: A Delphi Study on the Transformative Role of Artificial Intelligence in Higher Education

Meeta Joshi(✉) and Janvi Bhindi

Faculty of Management Studies, Marwadi University, Rajkot, India
meeta.joshi@marwadieducation.edu.in

Abstract. The study aims to investigate the ways in which AI is being utilized in HEIs and the possible impact that it may have on the future of education and teaching. To anticipate the various ways in which AI might be employed and the consequences it will have on HEIs, the research intends to collect the opinion of specialists using Delphi Methodology. The study's findings indicated that AI has the potential to significantly alter a variety of HEIs. Personalized learning experiences, AI based examinations and assessment, admission processes, and student support services are among the most significant applications that have been identified. Artificial Intelligence is the most effective factor for the cultivation of interpersonal skills, including collaboration, creativity and critical thinking, both for the students and the educators. The outcomes contribute to institutional leasers, legislators and educators' valuable insights as they formulate a strategy for HEIs that are influenced by AI.

Keywords: Artificial Intelligence · Delphi Technique · Higher Education · Student Support

1 Introduction

Higher Education Institutions have been cultivated for thousands of years, originating in ancient Greece and India. Ancient Greece was home to illustrious educational institutions like Plato's academy and Aristotle's lyceum, which significantly contributed to the advancement of knowledge and science. India has a long history of attracting scholars from around the globe to its renowned historical educational institutions like Takshashila and Nalanda, which has contributed to the country's rise to prominence as key hub for the dissemination of knowledge. The age of enlightenment further transformed universities into centers of scientific and intellectual inquiry (Lakshmi et al., 2023), catalyzing breakthroughs across various disciplines. The 19th & 20th Centuries bore witness to profound changes in higher education due to the industrial revolution and the emergence of modern educational systems. The epoch witnesses unprecedented global expansion of universities, offering an extensive spectrum of disciplines, encompassing science, engineering, social science, and humanities (Elimadi et al., 2024; Paton, 2018; SN, 2023).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 435–451, 2025.
https://doi.org/10.1007/978-3-031-86305-9_30

Emerging nearly seven decades ago, AI has embarked on an extraordinary evolution that continues to reverberate across the contemporary landscape. AI's profound and revolutionary impact spans industries, causing rapid disruptions in healthcare, aerospace, education, law, manufacturing, digital agriculture and transportation(Ali and Abdel-Haq, 2020; Bhutoria, 2022; PIDDUBTSEVA, 2024; Sabharwal et al., 2023). The pervasive influence of AI, capable of mimicking tasks once reserved solely for human cognition is remolding the facets of personal and professional existence. At its core, AI emerges as the convergence point for cutting-edge technologies, forging a symbiotic relationship with human intelligence. The emergence of AI-powered intelligent systems (Singh and Hiran, 2022; SN, 2023) manifests an uncanny prowess to "Think," "Comprehend," and "Act" mirroring human cognitive processes. The amalgamation of computer vision and audio processing equips AI with an understanding of its environment akin to human perception, while the tandem of Natural Language Processing(Ranga, 2024; Sabharwal et al., 2023; Sharma et al., 2022.a; Yadav et al., 2023.a) and Knowledge representation empowers AI to engage in nuanced comprehension and communication. Moreover, AI's toolbox includes the dexterity of Machine Learning and Expert Systems, enabling it to glean insights from data, make calculated decisions and execute intricate tasks with unparalleled precision (El-Adl, 2024a; Et al., 2021; Zmyzgova et al., 2020a). As AI's trajectory propels us towards a future that intertwines human and machine intelligence, its transformative influence only grows more compelling (AlDhaen, 2022).

The onset of the COVID-19 pandemic induced seismic shirts across global industries, yielding an unprecedented demand for innovative demand for innovative AI-driven educational solutions (AlDhaen, 2022; Ara Shaikh et al., 2022a; Cavus et al., 2021a; eldeeb et al., 2022; Et al., 2021; Nyanga, 2012; Photopoulos & Triantis, 2023a; Siddique and Chow, 2021) A survey conducted by the University Professional and continuing Education Association in 2021 illuminated a significant uptick in faculty members' confidence in online education underscoring the transformative influence of AI on pedagogical paradigm (Baskara, 2023; Creely and Blannin, 2023; Pillai et al., 2024). Among AI's myriad impacts, adaptive learning platforms stand out for their adeptness in tailoring content and pacing to individual students' needs. The incorporation of virtual reality simulations affords students immersive learning experiences (Elimadi et al., 2024; Mcintosh, 2005; Vargas-Murillo et al., 2023; Vargas-Murillo, Pari-Bedoya, et al., 2023), enabling the cultivation of real-world skills within controlled environments. Furthermore, the integration of AI-powered grading systems expedites assessment, granting educators more bandwidth to impart personalized instruction and feedback. AI's symbiotic relationship with higher education extends across diverse domains, encompassing machine learning, natural language processing, data mining, neural networks, and algorithms (Lovett et al., 2008).

2 Research Gap

Interestingly, a notable research gap persists in the underutilization of the Delphi technique as a tool of comprehending AI's transformative influence in HEIs. While numerous ai-related studies abound, few have tapped into the structured approach offered by the Delphi method to solicit expert consensus (Aithal and Aithal, 2023; Alnajim, 2023; Ara Shaikh et al., 2022b; Baskara, 2023a; Cavus et al., 2021b; El-Adl, 2024b; II, 2019; Karimi and Khawaja, 2023; Mallik and Gangopadhyay, 2023; Photopoulos and Triantis, 2023b; Sharma et al., 2022b; Yadav et al., 2023b; Zmyzgova et al., 2020b). The omission restricts the depth of comprehension concerning intricate AI integration challenges. By incorporating the Delphi technique into research endeavours, scholars could holistically assess AI's multifaceted implication uncovering subtleties that conventional methods may overlook. This methodological enhancement bears the potential to bridge existing gaps and contribute comprehensive insights for informed decision – making, thereby optimizing the role of AI in higher education.

Round 1 - AI Applications in Higher Education

Theme	Question
PL-AI (Personalized Learning with AI)	a. What are the most promising AI technologies for enabling personalized learning experiences in higher education?
	b. What are the potential benefits and challenges of implementing AI-driven personalized learning, and how can institutions address these challenges effectively?
SS-AI (AI for Student Support and Engagement)	a. How can AI-powered student support systems enhance engagement and academic success for students in higher education?
	b. What are the key factors to consider when developing AI-driven support tools to ensure they effectively cater to students' needs?
LA-AI (AI-Driven Learning Analytics)	a. How can AI-driven learning analytics contribute to enhancing the teaching and learning process in higher education?
	b. What concerns, if any, do you have about using AI for learning analytics, and what measures can be taken to address these concerns?
IM-AI (AI and Institutional Management)	a. Which areas of institutional management can benefit most from AI integration, and how can AI streamline administrative processes?
	b. What are the potential challenges and opportunities in adopting AI for institutional management, and how can these challenges be overcome?
AE-AI (AI and Accessibility in Education)	a. How can AI be utilized to create a more inclusive and accessible learning environment for students with disabilities in higher education?
	b. What steps can institutions take to ensure that AI-driven accessibility tools meet diverse needs and provide equal learning opportunities?

Round 1 – AI Application in Higher Education

Round 2 - Envisioning the AI-Driven Future	
Theme	Question
FT-AI (AI and the Future of Teaching)	a. What are your thoughts on the potential impact of AI on the role of educators, and how can educators embrace AI technologies effectively?
	b. In your opinion, what are the most significant challenges in integrating AI into teaching practices, and how can they be addressed?
VLE-AI (AI and Virtual Learning Environments)	a. How might AI-driven virtual learning environments revolutionize remote and online education, and what role can AI play in enhancing student engagement?
	b. What ethical considerations should be taken into account when deploying AI technologies in virtual learning environments?
HRI-AI (AI and Human-Robot Interaction)	a. What innovative applications of AI-powered robots can you envision for educational purposes, and how might they impact student learning experiences?
	b. What potential concerns or limitations should be considered in the implementation of human-robot interaction in educational settings?
OER-AI (AI and Open Educational Resources)	a. How can AI improve the discovery and recommendation of relevant Open Educational Resources (OER) for educators and students?
	b. What challenges may arise when using AI to generate or curate OER, and how can content quality and credibility be ensured?
SR-AI (AI and Student Retention)	a. How can AI be utilized to predict and address factors contributing to student attrition and improve overall student retention rates in higher education?
	b. What potential ethical concerns should be taken into account when using AI to monitor and support student retention efforts?

Round 2 – Envisioning the AI – Driven Future

Round 3 - Roadmap for Integration	
Theme	Question
PAA-AI (AI and Predictive Analytics for Admissions)	a. How can predictive analytics powered by AI be employed to make more informed and equitable admissions decisions?
	b. What steps should institutions take to ensure that AI-driven admissions processes are fair, transparent, and unbiased?
AE-AI (AI and Adaptive Examinations)	a. How can AI be effectively employed to create adaptive examination systems that cater to individual student abilities and learning trajectories?
	b. What challenges might arise during the implementation of AI-based adaptive testing, and how can these challenges be addressed?
TPD-AI (AI and Teacher Professional Development)	a. How can AI support personalized professional development opportunities for educators, and what resources or tools would be necessary for its successful implementation?
	b. What concerns or limitations might educators have regarding AI-driven professional development, and how can they be addressed?
CAD-AI (AI-Enabled Curriculum Design and Assessment)	a. How can AI assist educators in designing more effective curricula and assessments, and how can it encourage creativity and critical thinking among students?
	b. What potential drawbacks or misconceptions should be considered when incorporating AI into curriculum design and assessment processes?
DDDM-AI (AI and Data-Driven Decision Making)	a. How can AI-supported data analysis facilitate more informed decision-making processes in higher education institutions?
	b. What steps should institutions take to ensure that AI-driven decisions are transparent, fair, and accountable?

Round 3 – Roadmap for integration

3 Literature Review

The origins of AI and chatbots can be traced back to the 1950s when researchers began delving into the concept of artificial intelligence (Androutsopoulou et al., 2019; Rudolph et al., 2023a; Sandu and Gide, 2019a; Wang et al., 2023; Yang and Evans, 2019a). AI held the potential to be transformative technology in that era. Over recent years, significant advancements have taken place in AI, indicating its emergence as a technology poised to reshape human lives. Despite the global interest and the tangible reality of AI (Androutsopoulou et al., 2019; Baskara, 2023b; Wang, 2023; Yang and Evans, 2019a), the literature on this subject remains in its early stages. The educational realm has

become a focal point for the application of AI, garnering attention from scholars, educators, governments and researchers. AI offers vast opportunities to enhance teaching and learning. In fact, it has ushered in a more efficient, adaptable and effective educational model compared to traditional schooling. Within the education sector, AI can interact with, assisting, and guiding students in their self-exploration education (SEE) through (Berezinska, 2023; La Torre et al., 2023; Naidu, 2019; Yadav et al., 2023; Zmyzgova et al., 2020c) (SDL) or Hauntological techniques (Heilala et al., 2023), akin to contemporary, semi-automatic learning systems. The utilization of AI in higher education constitutes a burgeoning field with multifaceted applications and implications (Du et al., 2023; Shamkuwar et al., 2023). AI technologies in higher education have the potential to enrich the learning experience, provide student support (Al-Maskari et al., 2024; Antony and Ramnath, 2023), streamline enrollment management, and enhance academic integrity (Peterson and Keller, 1984; Slowey and Kozina, 2015; Useem, 1998). However, certain aspects of best practice, such as addressing collusion and cheating, might not be fully integrated. (Altbach et al., 2009) spotlighted successful AI technology application across three main domains of college operations: enhancing student learning experience, providing student support, and optimizing enrolment management. AI bears the potential to tackle significant education challenges (Abulibdeh et al., 2024; Sok and Heng, 2023), overhaul teaching and learning methodologies, and even bring about structural changes in university administrative services. In a similar vein, (Bucea-Manea-ţoniş et al., 2022) conducted an exploratory survey involving 139 teachers from Romania and Serbia. The survey demonstrates the willingness of teachers to adopt AI in HEIs with a focus on seizing opportunities rather than fixating on challenges. Caplan et al., 2023 proposed that the integration of AI in HEIs can be help mitigate some of the challenges faced by the sector, such as the influx of students and the demand for personalised learning. Overall, AI has the potential to revolutionize higher education, but ethical considerations and evidence-based recommendations should guide its implementation (Abulibdeh et al., 2024c; Jeyaraman et al., 2023; MOHSIN, 2020). The introduction of AI in mentoring processes raises ethical concerns, such as data protection, non-discrimination, integrity of data, safety, security, confidentiality, bias, trust and transparency and cyber security plays a significant role in the successful and effective adoption of AI in HLIs, as universities and colleges are common targets of cyber attackers (Memarian and Doleck, 2023).

4 Research Methods

The Delphi procedure, a systematic and proficient technique, was utilized to collect and synthesize the informed opinions of domain specialists on intricate situations, particularly in areas characterized by doubt or disagreement. This study employed a methodical and systematic approach by convening a broad and esteemed panel of twelve experts from the domains of AI, education, administration, and related topics. Their collective expertise was the foundation of this research, yielding substantial analytical breakthroughs. A primary objective in this procedure was to uphold impartiality and reduce prejudice, accomplished through a meticulously crafted methodology that guaranteed the study's neutrality. The researchers followed a rigorous protocol, performing three consecutive

rounds of interviews, each increasingly enhancing the results. The initial round concentrated on the implementation of AI in Higher Educational Institutions (HEIs). The subjects investigated were categorized into coded themes, facilitating a comprehensive analysis of AI's function in education. The data collection process involved face-to-face interviews, effectively capturing the intricacies of expert viewpoints. Table 1 presents a summary of expert assessments regarding the potential of AI technology to improve individualized learning experiences in higher education institutions. This table summarizes expert assumptions, consensus levels, and diverse viewpoints on several AI-driven technologies. The aim of the initial round was to collect pertinent insights from specialists regarding the fundamental AI-driven techniques. These assessments establish a crucial basis for the formulation of a strategic framework that influences individualized learning experiences in higher education institutions. The data study uncovered significant insights regarding AI's prospective uses in higher education, especially with customized learning experiences.

The specialists assessed many theme domains, including:

1. Personalized Learning with AI (PL-AI): The function of AI in tailoring educational materials to meet the specific requirements of individual pupils.
2. Student Support and Engagement with AI (SSI-AI): The potential of AI to augment student support frameworks and engagement methodologies.
3. AI-Enhanced Learning Analytics (LA-AI): The application of AI to evaluate student performance and educational results.
4. AI in Institutional Management (IM-AI): The role of AI in optimizing administrative procedures within Higher Education Institutions (HEIs).
5. AI and Educational Accessibility (AE-AI): The potential of AI to enhance educational access for various student demographics, especially individuals with special needs.

These evaluations offer profound insights into the potential of AI technology, as well as the problems and opportunities inherent in each theme area. The debates indicated that although AI can greatly improve personalized learning and student engagement, there are considerable concerns around ethical considerations, data protection, and the necessity for sufficient infrastructure in higher education institutions to facilitate AI-driven systems.

Furthermore, in later rounds, experts honed their viewpoints, resulting in an increased consensus regarding the strategic ramifications of AI in education. This process enabled the identification of critical obstacles, including the integration of AI into current systems, guaranteeing ethical utilization, and resolving potential inequities in access to AI technologies among institutions.

Strategy for Evaluation

All consent through the experts were considered before evaluating the data received from the Experts. To employ the linguistic terminology and the evaluation, like high agreement, moderate agreement and poor agreement, the assessment was converted into fuzzy numbers to enable quantification.

- A strong consensus was represented as a fuzzy number with values approaching 1 (e.g., (0.8, 0.9, 1)).

Table 1. Round 1: Expert assessment on Promising AI Technologies in HEIs

Round 1										
Theme	**Questions**	**Code**	**Experts' Responses**	**Mean**	**Standard Deviation**	**Linguistic Rating**	**Fuzzy Number**	**Fuzzy Delphi Score (A) ≥ 0.5**	**(%) Expert Consensus ≥ 80%**	**Expert Consensus**
PL-AI (Personalized Learning with AI)	a. What are the most promising AI technologies for enabling personalized learning experiences in higher education?	PL1	By analyzing historical admissions data and student performance metrics, AI can predict the likelihood of an applicant's success in a particular program.	4.15	0.58	High agreement	(0.8, 1, 1)	0.933	89%	Accept
		PL2	AI-driven predictive models can identify and prioritize key attributes that contribute to student success, enabling more data-driven admissions decisions.	4.1	0.61	Medium agreement	(0.5, 0.8, 0.4)	0.789	67%	Reject
		PL3	AI-powered predictive analytics can help institutions implement targeted outreach and support programs for students at risk of not completing their studies.	4.25	0.53	Low agreement	(0.3, 0.6, 0.4)	0.743	78%	Reject
	b. What are the potential benefits and challenges of implementing AI-driven personalized learning, and how can institutions address these challenges effectively?	PL4	Implementing a comprehensive audit trail to track and explain the decision-making process of AI algorithms for admissions.	4.18	0.63	Low agreement	(0.3, 0.5, 0.7)	0.478	50%	Reject
		PL5	Regularly evaluating and updating the AI models to address any biases and ensure fairness and inclusivity in the admissions process.	4.13	0.66	Medium agreement	(0.4, 0.6, 0.8)	0.567	63%	Reject
		PL6	Engaging diverse stakeholders, including representatives from underrepresented groups, to provide input and oversight on the AI-driven admissions process.	4.22	0.55	Medium agreement	(0.6, 0.5, 0.5)	0.541	75%	Reject
SS-AI (AI for Student Support and Engagement)	a. How can AI-powered student support systems enhance engagement and academic success for students in higher education?	SS1	AI-driven support systems can provide personalized recommendations for resources and activities, promoting student engagement and learning outcomes.	4.28	0.52	High agreement	(0.9, 1, 1)	0.923	89%	Accept
		SS2	AI chatbots can offer 24/7 assistance, answering student queries promptly, enhancing student support and reducing response time.	4.13	0.61	High agreement	(0.8, 0.9, 1)	0.89	85%	Accept
		SS3	AI analytics can identify students at risk of academic challenges, enabling timely interventions and personalized support for their success.	4.25	0.53	High agreement	(0.8, 0.9, 1)	0.9	85%	Accept
	b. What are the key factors to consider when developing AI-driven support tools to ensure they effectively cater to students' needs?	SS4	Designing user-friendly interfaces and interactive platforms to enhance accessibility and ease of use for diverse student populations.	4.15	0.58	Medium agreement	(0.5, 0.7, 0.4)	0.517	67%	Reject
		SS5	Ensuring data privacy and security to build trust among students and protect sensitive information shared within the support tools.	4.1	0.63	High agreement	(0.9, 1, 1)	0.998	93%	Accept
		SS6	Involving students in the development process to gather feedback and insights, ensuring that the AI-driven tools align with their actual needs and preferences.	4.22	0.55	Low agreement	(0.4, 0.6, 0.8)	0.321	60%	Reject
LA-AI (AI-Driven Learning Analytics)	a. How can AI-driven learning analytics contribute to enhancing the teaching and learning process in higher education?	LA1	AI-driven learning analytics can provide real-time insights into student progress and performance, enabling timely interventions and personalized support.	4.32	0.47	High agreement	(0.7, 0.9, 1)	0.872	78%	Reject
		LA2	AI can identify patterns in student learning behavior, helping instructors tailor their teaching approaches to better meet individual learning needs.	4.17	0.52	Medium agreement	(0.5, 0.8, 1)	0.51	67%	Reject
		LA3	AI analytics can assess the effectiveness of instructional materials and teaching methods, leading to continuous improvement of teaching practices.	4.25	0.49	High agreement	(0.8, 0.9, 1)	0.945	85%	Accept
	b. What concerns, if any, do you have about using AI for learning analytics, and what measures can be taken to address these concerns?	LA4	Ensuring data privacy and confidentiality of student information when using AI for learning analytics, and implementing robust data protection measures.	4.1	0.57	High agreement	(0.9, 1, 1)	0.911	93%	Accept
		LA5	Addressing potential biases in AI algorithms to prevent any adverse impact on underrepresented student groups and ensure equitable learning outcomes.	4.08	0.58	High agreement	(0.8, 1, 1)	0.896	89%	Accept
		LA6	Providing transparent explanations of AI-generated insights to instructors and students to foster trust and understanding of AI-driven learning analytics.	4.2	0.51	Medium agreement	(0.6, 0.8, 1)	0.673	67%	Reject
IM-AI (AI and Institutional Management)	a. Which areas of institutional management can benefit most from AI integration, and how can AI streamline administrative processes?	IM1	AI integration can greatly benefit student admissions and enrollment management. By automating the application screening process, AI can efficiently analyze applicant data, academic records, and other relevant information to identify suitable candidates, reducing manual effort and speeding up decision-making.	4.5	0.32	High agreement	(0.8, 0.9, 1)	0.897	85%	Accept
		IM2	AI can revolutionize financial management in institutions by optimizing budget allocation and expenditure tracking. Through AI-powered predictive analytics, institutions can make data-driven financial decisions, identify cost-saving opportunities, and ensure efficient resource utilization.	4.3	0.41	High agreement	(0.7, 1, 1)	0.888	89%	Accept
		IM3	AI-driven analytics in academic performance assessment can lead to more effective student support and intervention strategies. By analyzing student learning patterns and early warning signs, institutions can provide personalized academic assistance, enhancing student success and retention rates.	4.4	0.33	High agreement	(0.8, 1, 1)	0.861	89%	Accept
	b. What are the potential challenges and opportunities in adopting AI for institutional management, and how can these challenges be overcome?	IM4	One potential challenge in adopting AI for institutional management is the initial cost and resource investment required for implementation. Institutions may need to allocate significant funds and time to procure AI technology, train staff, and integrate AI systems with existing processes.	3.8	0.41	Low agreement	(0.3, 0.6, 0.9)	0.331	60%	Reject
		IM5	Another challenge lies in the ethical and privacy concerns related to AI-driven decision-making. Institutions must ensure transparency and fairness in AI algorithms to avoid biased outcomes and protect sensitive student data from unauthorized access.	4.1	0.44	Medium agreement	(0.6, 0.8, 1)	0.431	67%	Reject
		IM6	Despite challenges, AI presents numerous opportunities to enhance institutional management. AI can improve operational efficiency by automating repetitive tasks, freeing up staff to focus on more strategic initiatives. Additionally, AI-driven insights can enable data-informed decision-making, leading to better resource allocation and student support.	4.3	0.45	Medium agreement	(0.7, 0.9, 1)	0.521	78%	Reject

IM-AI (AI and Institutional Management)	a. Which areas of institutional management can benefit most from AI integration, and how can AI streamline administrative processes?	IM1	AI integration can greatly benefit student admissions and enrollment management. By automating the application screening process, AI can efficiently analyze applicant data, academic records, and other relevant information to identify suitable candidates, reducing manual effort and speeding up decision-making.	4.5	0.32	High agreement	(0.8, 0.9, 1)	0.897	85%	Accept
		IM2	AI can revolutionize financial management in institutions by optimizing budget allocation and expenditure tracking. Through AI-powered predictive analytics, institutions can make data-driven financial decisions, identify cost-saving opportunities, and ensure efficient resource utilization.	4.3	0.41	High agreement	(0.7, 1, 1)	0.888	89%	Accept
		IM3	AI-driven analytics in academic performance assessment can lead to more effective student support and intervention strategies. By analyzing student learning patterns and early warning signs, institutions can provide personalized academic assistance, enhancing student success and retention rates.	4.4	0.33	High agreement	(0.8, 1, 1)	0.861	89%	Accept
	b. What are the potential challenges and opportunities in adopting AI for institutional management, and how can these challenges be overcome?	IM4	One potential challenge in adopting AI for institutional management is the initial cost and resource investment required for implementation. Institutions may need to allocate significant funds and time to procure AI technology, train staff, and integrate AI systems with existing processes.	3.8	0.41	Low agreement	(0.3, 0.6, 0.9)	0.331	60%	Reject
		IM5	Another challenge lies in the ethical and privacy concerns related to AI-driven decision-making. Institutions must ensure transparency and fairness in AI algorithms to avoid biased outcomes and protect sensitive student data from unauthorized access.	4.1	0.44	Medium agreement	(0.6, 0.8, 1)	0.431	67%	Reject
		IM6	Despite challenges, AI presents numerous opportunities to enhance institutional management. AI can improve operational efficiency by automating repetitive tasks, freeing up staff to focus on more strategic initiatives. Additionally, AI-driven insights can enable data-informed decision-making, leading to better resource allocation and student support.	4.3	0.45	Medium agreement	(0.7, 0.9, 1)	0.521	78%	Reject

- Moderate agreement was indicated by results near the midway (e.g., 0.4, 0.5, 0.6).
- Minimal agreement was indicated by numbers approaching 0 (e.g., (0, 0.1, 0.2)).

Following the conversion of all linguistic ratings into the fuzzy numbers, the fuzzy scores were calculated by aggregating the fuzzy numbers supplied by each expert. The predominant technique employed was the fuzzy average, which entailed aggregating the fuzzy values for a certain criterion from all experts and dividing by the total number of experts.

Let *fi* denote the fuzzy number assigned by the i-th expert, and let *N* represent the total number of experts. This yielded a unified fuzzy score for each criterion or decision variable.

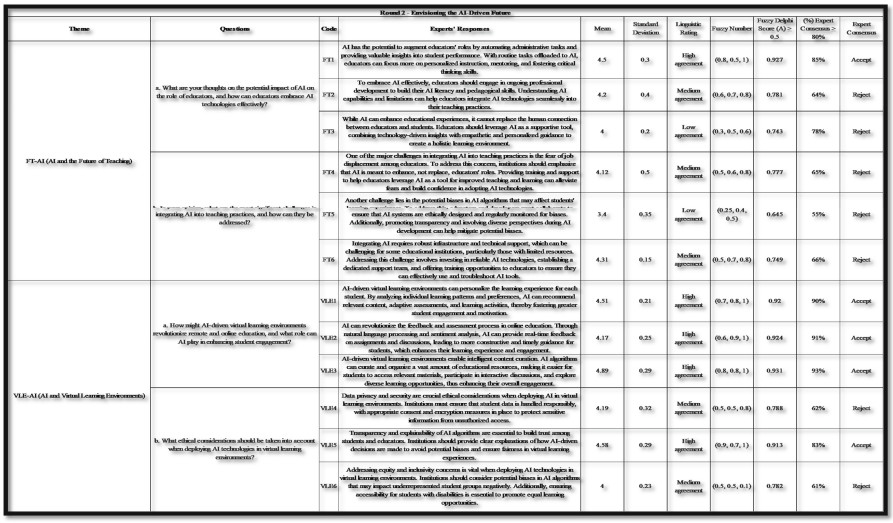
The fuzzy Delphi score was obtained by additional processing of the aggregated fuzzy scores to indicate the degree of consensus. Following each round of expert assessments, modifications were implemented depending on the feedback received, and revised fuzzy scores were calculated. The Delphi score generally integrates the fuzzy score with a metric of variability or consensus across experts. A high level of consensus produced more stable and convergent fuzzy Delphi scores, whereas diverging perspectives resulted in wider ranges of fuzzy numbers. Expert commentary was essential in elucidating the ambiguous scores. Following each evaluation session, experts were prompted to offer qualitative insights or elucidations. The comments were examined for patterns, and rationales for divergent perspectives were evaluated. In certain instances, this feedback was utilized to amend the evaluation criteria or to enhance the focal areas for future iterations of the Delphi process.

At last the expert consensus was determined by the level of convergence in the fuzzy scores across several rounds. A robust consensus was evident as the fuzzy numbers exhibited increased concentration, characterized by reduced intervals between the minimum and maximum values. The consensus level was articulated using standard deviation or coefficient of variance in fuzzy scores. In instances with low consensus, further feedback rounds were conducted to reconcile differing viewpoints and steer experts towards enhanced agreement.

Table 2 delineates an advanced perspective on the potential implications of AI concerning the HEIs. The second round was dedicated to the development of AI-Driven technologies in future HEIs, encompassing Future teaching with AI (FT-AI), Open educational resources and AI (OER-AI), Human interaction with AI (HRI-AI), Open educational resources with AI(OER-AI), and student retention with AI(SR-AI). The comprehensive analysis presented in Table 2 encapsulated the consensus and dissent among experts.

Table 3 displays the findings of the third round with experts' where there is integration of AI technology in the HEIs where discussed. Experts discussed the roadmap that precisely outlines actionable steps, key considerations and the potential challenges in the integrations process. The table further represents the overview of the expert' sentiments, levels of agreement, and future outlooks, providing valuable information on the strategic path for maximizing the potential of AI to transform and enhance the educational experience in the HEIs. The major aim of this round 3 is to develop AI main domain in education domain. The themes encompassed are AI in admission process (PAA-AI), Adaptive examinations & Assessment (APE-AI), teacher professional development with

Table 2. Round 2 – Experts' assessment on development of AI-driven technologies in Future HEIs



Round 2 - Envisioning the AI-Driven Future										
Theme	Questions	Code	Experts' Responses	Mean	Standard Deviation	Linguistic Rating	Fuzzy Number	Fuzzy Delphi Score (A) > 0.5	Obj Expert. Consensus > 80%	Expert Consensus
FT-AI (AI and the Future of Teaching)	a. What are your thoughts on the potential impact of AI on the role of educators, and how can educators embrace AI technologies effectively?	FT1	AI has the potential to augment educators' roles by automating administrative tasks and providing valuable insights into student performance. With routine tasks offloaded to AI, educators can focus more on personalized instruction, mentoring, and fostering critical thinking skills.	4.5	0.3	High agreement	(0.8, 0.5, 1)	0.927	85%	Accept
		FT2	To embrace AI effectively, educators should engage in ongoing professional development to build their AI literacy and pedagogical skills. Understanding AI capabilities and limitations can help educators integrate AI technologies seamlessly into their teaching practices.	4.2	0.4	Medium agreement	(0.6, 0.7, 0.8)	0.781	64%	Reject
		FT3	While AI can enhance educational experiences, it cannot replace the human connection between educators and students. Educators should leverage AI as a supportive tool, combining technology-driven insights with empathetic and personalized guidance to create a holistic learning environment.	4	0.2	Low agreement	(0.3, 0.5, 0.6)	0.743	78%	Reject
	integrating AI into teaching practices, and how can they be addressed?	FT4	One of the major challenges in integrating AI into teaching practices is the fear of job displacement among educators. To address this concern, institutions should emphasize that AI is meant to enhance, not replace, educators' roles. Providing training and support to help educators leverage AI as a tool for improved teaching and learning can alleviate fears and build confidence in adopting AI technologies.	4.12	0.5	Medium agreement	(0.5, 0.6, 0.8)	0.777	65%	Reject
		FT5	Another challenge lies in the potential biases in AI algorithms that may affect students' ensure that AI systems are ethically designed and regularly monitored for biases. Additionally, promoting transparency and involving diverse perspectives during AI development can help mitigate potential biases.	3.4	0.35	Low agreement	(0.25, 0.4, 0.5)	0.645	55%	Reject
		FT6	Integrating AI requires robust infrastructure and technical support, which can be challenging for some educational institutions, particularly those with limited resources. Addressing this challenge involves investing in reliable AI technologies, establishing a dedicated support team, and offering training opportunities to educators to ensure they can effectively use and troubleshoot AI tools.	4.31	0.15	Medium agreement	(0.5, 0.7, 0.8)	0.749	66%	Reject
VLE-AI (AI and Virtual Learning Environments)	a. How might AI-driven virtual learning environments revolutionize remote and online education, and what role can AI play in enhancing student engagement?	VLE1	AI-driven virtual learning environments can personalize the learning experience for each student. By analyzing individual learning patterns and preferences, AI can recommend relevant content, adaptive assessments, and learning activities, thereby fostering greater student engagement and motivation.	4.51	0.21	High agreement	(0.7, 0.8, 1)	0.92	90%	Accept
		VLE2	AI can revolutionize the feedback and assessment process in online education. Through natural language processing and sentiment analysis, AI can provide real-time feedback on assignments and discussions, leading to more constructive and timely guidance for students, which enhances their learning experience and engagement.	4.17	0.23	High agreement	(0.6, 0.9, 1)	0.924	91%	Accept
		VLE3	AI-driven virtual learning environments enable intelligent content curation. AI algorithms can curate and organize a vast amount of educational resources, making it easier for students to access relevant materials, participate in interactive discussions, and explore diverse learning opportunities, thus enhancing their overall engagement.	4.89	0.29	High agreement	(0.8, 0.8, 1)	0.931	93%	Accept
	b. What ethical considerations should be taken into account when deploying AI technologies in virtual learning environments?	VLE4	Data privacy and security are crucial ethical considerations when deploying AI in virtual learning environments. Institutions must ensure that student data is handled responsibly, with appropriate consent and encryption measures in place to protect sensitive information from unauthorized access.	4.19	0.32	Medium agreement	(0.5, 0.5, 0.8)	0.788	62%	Reject
		VLE5	Transparency and explainability of AI algorithms are essential to build trust among students and educators. Institutions should provide clear explanations of how AI-driven decisions are made to avoid potential biases and ensure fairness in virtual learning experiences.	4.58	0.29	High agreement	(0.9, 0.7, 1)	0.913	83%	Accept
		VLE6	Addressing equity and inclusivity concerns is vital when deploying AI technologies in virtual learning environments. Institutions should consider potential biases in AI algorithms that may impact underrepresented student groups negatively. Additionally, ensuring accessibility for students with disabilities is essential to promote equal learning opportunities.	4	0.23	Medium agreement	(0.5, 0.5, 0.1)	0.782	61%	Reject

AI (TPD-AI), AI-enabled curriculum design and assessment(CAD-AI), and data-driven decision-making (DDM-AI). These themes emerge from the perspective of experts' opinions and evaluations,

5 Analyzing the Journey: AI Integration in Higher Education

Round one, examine the possibilities of AI in HEIs and how AI technologies can further enhance the HEIs in various domains of Higher education. The experts have given their assessment of the various AI application and how these applications can be dominant in AI – driven technologies for the various themes of the HEIS. These themes are emphasizing on personalized learning experiences, student support, admission process, administrative procedures and examinations and assessments. However, there are skeptical opinions of the experts' concerning ethical dilemma, cyber security, transparency, trust and biases in the usage of the AI in HEIs. Round second, our attention was towards the possible chances given by upcoming ai-driven developments in Higher education. Among several concepts, a consensus developed in this field of progressive views. Professionals all agree that AI has transforming ability to change educational methods. It relates to the concepts of boosting teachers' responsibilities, personalized learning environment, and strengthen evaluation systems. Still among this harmonic concord, there were some differences. Different points of view have highlighted issues including ethical conundrums, the predominance of AI over human contact, and the accidental support of standardized testing, in Round third, the expedition peaked, and the major focus shifted to developing a comprehensive strategy for the flawless integration of AI in HEIs. Cooperative efforts have cleared the path for the application of AI in significant spheres

Table 3. Round 3: Roadmap for integration of AI in HEIs

Round 3 - Roadmap for Integration										
Theme	Questions	Code	Experts' Responses	Mean	Standard Deviation	Linguistic Rating	Fuzzy Number	Fuzzy Delphi Score (A) ≥ 0.5	(%) Expert Consensus ≥ 80%	Expert Consensus
PAAI-AI (AI and Predictive Analytics for Admissions)	a. How can predictive analytics powered by AI be employed to make more informed and equitable admissions decisions?	PAA1	Predictive analytics can leverage historical data and machine learning algorithms to identify patterns and factors that contribute to successful student outcomes.	4.6	0.3	Low agreement	(0.4, 0.2, 1)	0.326	65%	Reject
		PAA2	By analyzing various data points, such as academic performance, extracurricular activities, and socio-economic background, AI can help institutions assess applicants holistically.	4.3	0.3	Medium agreement	(0.5, 0.8, 0.4)	0.789	67%	Reject
		PAA3	Employing predictive analytics can enable universities to make more informed and unbiased decisions, promoting equity and diversity in the admissions process.	4.1	0.21	High agreement	(0.9, 0.7, 0.6)	0.943	92%	Accept
	b. What steps should institutions take to ensure that AI-driven admissions processes are fair, transparent, and unbiased?	PAA4	Establish clear objectives and guidelines for the AI-driven admissions process, prioritizing fairness and equity in student selection. Involve a diverse group of stakeholders, including ethicists, educators, and experts in AI, to oversee and validate the AI-driven admissions system for transparency and unbiased decision-making.	4.2	0.3	High agreement	(0.6, 0.9, 1)	0.924	91%	Accept
		PAA5	Regularly assess and monitor the AI algorithms to identify and rectify any biases that may arise from the data or the models themselves.	4.4	0.33	Low agreement	(0.5, 0.8, 1)	0.322	67%	Reject
		PAA6	Involve a diverse group of stakeholders, including ethicists, educators, and experts in AI, to oversee and validate the AI-driven admissions system for transparency and unbiased decision-making.	4.35	0.23	Medium agreement	(0.4, 0.6, 0.8)	0.567	63%	Reject
APE-AI (AI and Adaptive Examinations)	a. How can AI be effectively employed to create adaptive examination systems that cater to individual student abilities and learning trajectories?	APE1	AI can analyze student performance data and learning patterns to create personalized assessment paths.	4.5	0.23	High agreement	(0.7, 0.8, 1)	0.92	90%	Accept
		APE2	Machine learning algorithms can adapt the difficulty and content of questions based on individual strengths and weaknesses.	4.7	0.23	High agreement	(0.6, 0.9, 1)	0.924	91%	Accept
		APE3	Real-time feedback from AI-driven systems can help students identify areas of improvement and enhance their learning experience.	4.8	0.21	Medium agreement	(0.5, 0.5, 0.8)	0.781	62%	Reject
	b. What challenges might arise during the implementation of AI-based adaptive testing, and how can these challenges be addressed?	APE4	Data privacy concerns may arise as AI systems require access to sensitive student information. Implement robust privacy measures and obtain consent from students and parents.	4.9	0.3	Medium agreement	(0.3, 0.5, 0.8)	0.788	69%	Reject
		APE5	Ensuring that AI algorithms are regularly updated and validated to prevent bias and maintain accuracy.	4.05	0.27	Medium agreement	(0.5, 0.5, 0.1)	0.782	61%	Reject
		APE6	Addressing technical issues and ensuring the system's accessibility to all students, including those with limited access to technology.	4.33	0.29	Medium agreement	(0.4, 0.3, 0.8)	0.799	79%	Reject
TPD-AI (AI and Teacher Professional Development)	a. How can AI support personalized professional development opportunities for educators, and what resources or tools would be necessary for its successful implementation?	TPD1	AI can analyze educators' teaching styles and strengths to recommend personalized training modules.	4.22	0.3	High agreement	(0.8, 0.81, 1)	0.934	91%	Accept
		TPD2	Natural Language Processing can provide feedback on communication skills.	4.63	0.56	High agreement	(0.7, 0.75, 1)	0.938	93%	Accept
		TPD3	AI-driven platforms with access to online courses, workshops, and peer networks can be essential resources for educators' growth.	4.42	0.45	Medium agreement	(0.5, 0.8, 0.4)	0.789	67%	Reject
	b. What concerns or limitations might educators have regarding AI-driven professional development, and how can they be addressed?	TPD4	Fear of AI replacing human interaction. Emphasize that AI is meant to enhance, not replace, human teaching skills.	4.27	0.2	Medium agreement	(0.5, 0.2, 0.2)	0.794	59%	Reject
		TPD5	Lack of trust in AI recommendations. Provide transparent explanations of AI algorithms and allow educators to have control over their professional development choices.	4.11	0.21	Medium agreement	(0.5, 0.5, 0.1)	0.72	67%	Reject
		TPD6	Address privacy concerns and ensure that data used for professional development is secure and anonymized.	4.49	0.23	High agreement	(0.9, 0.84, 1)	0.708	91%	Accept
CAD-AI (AI-Enabled Curriculum Design and Assessment)	a. How can AI assist educators in designing more effective curricula and assessments, and how can it encourage creativity and critical thinking among students?	CAD1	AI can analyze student performance data to identify areas where curricula need improvement.	4.52	0.57	Low agreement	(0.2, 0.3, 0.5)	0.68	54%	Reject
		CAD2	Natural Language Processing can assess written assignments for creativity and critical thinking.	4.47	0.66	Low agreement	(0.2, 0.3, 0.4)	0.648	51%	Reject
		CAD3	AI can offer personalized learning experiences, allowing students to explore diverse subjects and develop critical thinking skills.	4.61	0.59	High agreement	(0.8, 0.81, 1)	0.934	80%	Accept
	b. What potential drawbacks or misconceptions should be considered when incorporating AI into curriculum design and assessment processes?	CAD4	Over-reliance on AI can diminish human creativity and judgment. Emphasize the role of educators in the decision-making process.	4.16	0.43	High agreement	(0.7, 1, 1)	0.888	89%	Accept
		CAD5	AI algorithms might not fully understand the context or complexity of certain subjects, leading to inaccurate assessments.	4.02	0.17	Low agreement	(0.3, 0.3, 0.5)	0.68	49%	Reject
		CAD6	Address concerns about AI reinforcing standardized testing and neglecting other important aspects of education.	4.28	0.21	Medium agreement	(0.2, 0.5, 0.8)	0.71	77%	Reject
DDDM-AI (AI and Data-Driven Decision Making)	a. How can AI-supported data analysis facilitate more informed decision-making processes in higher education institutions?	DDAM1	AI can process large datasets to identify trends, patterns, and insights for strategic planning.	4.15	0.22	High agreement	(0.7, 0.52, 0.8)	0.73	89%	Accept
		DDAM2	Predictive analytics can assist in student enrollment and retention predictions.	4.67	0.52	High agreement	(0.9, 0.98, 1)	0.928	90%	Accept
		DDAM3	AI-driven data analysis can optimize resource allocation and budgeting in educational institutions.	4.45	0.54	High agreement	(0.9, 0.85, 1)	0.923	88%	Accept
	b. What steps should institutions take to ensure that AI-driven decisions are transparent, fair, and accountable?	DDAM4	Provide clear guidelines and policies on how AI is utilized in decision-making processes.	4.25	0.3	Medium agreement	(0.4, 0.3, 0.18)	0.789	77%	Reject
		DDAM5	Regularly audit and validate AI algorithms to ensure they are free from bias and errors.	4.12	0.2	Medium agreement	(0.4, 0.45, 0.8)	0.778	79%	Reject
		DDAM6	Involve diverse stakeholders, including students, educators, and experts, in the development and oversight of AI systems to enhance transparency and fairness.	4.01	0.11	Low agreement	(0.2, 0.3, 0.4)	0.648	79%	Reject

like admissions, assessments, professional growth, curriculum design and decision making, among these arguments, there were wary voices calling through the hallways. The research addresses prejudices, safeguarding privacy using a human-centered approach, avoiding too much reliance on AI. This stage required the sophisticated coordination required to negotiate the terrain of AI. Overall, the development over these 3 rounds of research revealed a wide spectrum of opinions on the integration of AI with HEIs.

6 Limitations of the Study

Regardless of this research future possibility, this research does hold significant limitations that need to be consideration. The Delphi technique is having organised methodology, though it adds bias by including experts who may hold notions for the AI based technology. This study may unintentionally assume a linear trajectory of AI improvement,

so underestimating the intricate and unpredictable nature of technology advancement, potentially impacting the precision of its forecasts. Due to challenges and complex variations of different educational systems, this is a challenge of generalizability of findings regarding the integration of AI in HEIs.

7 Synthesis as a Future Direction

Exploring the Advanced Education environment enhanced by AI: By combining the diverse insights obtained from all three rounds of analysis, there is a roadmap which can be outlined for seamlessly integrating AI in to HEIs. This roadmap supports in transforming education into a vibrant, student centric system that combines innovation with ethical and systematic pedagogical reviews for better education system. Balancing enthusiasm with Ethical Awareness: The juxtaposition of optimism and caution across all rounds highlights the delicate equilibrium education must strive for. While Round 2 embraced the transformative potential of AI, round 3's rejections underscore the necessity of ethical vigilance (Baradziej, 2023b; Jones and Masika, 2021b; Qian, 2023b; Sarwar et al., 2024c; Yu et al., 2023c). Institutions must heed these warnings as a beacon, ensuring that AI's implementation is guided by responsible frameworks. Transparency, fairness and privacy safeguarded will be the cornerstones to constructing an AI-Driven future that is both innovative and morally principled (González-Calatayud, et al., 2021a; Li et al., 2021b; Vial et al., 2023c; Zhang et al., 2023b). Empowering Stakeholder Collaboration: The journey across rounds underscores the importance of diverse voices. Collaboration among educators, technologists, policymakers and learners is essential. The divergent perspectives revealed in our study emphasis that a holistic approach, one that values human expertise and experience, must guide AI's evolution in education (Baradziej, 2023c; Camaréna, 2022c; Dautov et al., 2023c; Farazouli et al., 2024; Yu et al., 2023d). Experts' diverse viewpoints provide a reminder that while technology may be cutting-edge, its efficacy is amplified when harnessed collaboratively.

From Assumptions to Actionable Insights: the synthesis of these rounds urges institutions to translate insights into actionable strategies. Round 1's skepticism calls for thought impact assessments, ensuring that the integration of AI does not inadvertently sideline ethics, data privacy, or socio-economic concerns (Gašević et al., 2023a; González-Calatayud, et al., 2021b; Kalyani, 2023b). Round 2's consensus should fuel the development of personalized learning experiences and intelligent assessments that nurture critical thinking and creativity. Round 3's roadmap should be embraced as a blueprint, outlining steps to ensure the equitable use of AI, its transparent deployment and its validation as an instrument of positive change (Ahmad et al., 2019; Gašević et al., 2023b; Hameed, 2020; Zhang et al., 2023c). A Transformational Journey, not a Destination: the journey through these rounds underscores that I's integration is not a destination but a continuous evolution. As AI Technology evolves, so should the strategies and policies be surrounding its application. Institutions must commit to ongoing research, development and refinement to ensure that AI-driven advancements remain aligned with educational values and societal needs (Fernández-Gutiérrez et al., 2023; Lainjo and Tmouche, 2024; Rudolph et al., 2023b; Sandu and Gide, 2019b; Yang and Evans, 2019b). The transformative potential of AI can only be fully harnessed through an unwavering commitment to adaptation, iteration, and responsible stewardship.

8 Conclusion: Mapping Out a Revolutionary Path for AI Incorporation in Higher Education Institutions

This conclusion encapsulates insights obtained from all different rounds of the analysis which provide a detailed overview of dynamics associated with the integration of AI in higher education also how ethically and with the transparency the integration is possible can also be conceptualise. With the insights of all stakeholders and underscoring the strong technical foundations and progressive development, the integration of AI in higher education can be made possible. This summary also adds valuing diversity, confronting diverse problems, promising AI as a very strong catalyst for inclusivity and empowering the education system. This analysis will also be a guide for the policymakers and the educators to adopt practical measures that successfully utilize AI technologies guaranteeing equitable access and opportunity for all learners. At last, the strategic integration of AI in HEIs has the potential to transform educational practices, necessitating a unified approach to address ethical issues, promote collaboration among stakeholders and perpetually enhance methodologies to adapt to charging educational needs.

Annex – 1

Components and Interpretation Guide for Expert Ratings:

Question and Code: This section refers to the specific question being asked and is given a unique code (in this case, "PL1" for the first question).

Experts' Responses: Each row under the "Experts' Responses" column provides a brief description of an AI technology that is considered promising for personalized learning experiences in higher education.

Mean and Standard Deviation: The "Mean" is the average rating given by the experts to the specific AI technology, indicating their level of agreement on its potential. The "Standard Deviation" measures the extent of agreement or disagreement among the experts' ratings. A lower standard deviation suggests more consensus among experts.

Linguistic Rating: This column provides a qualitative description of the level of agreement among the experts. In this case, "High agreement," "Medium agreement," and "Low agreement" indicate the level of consensus on the respective technologies.

We utilized linguistic ratings (Low, Medium, or High) for experts' responses, which were then converted into numerical values (0, 0.5, or 1) using the following formula:

Numerical Value (NV) = (Linguistic Rating - Low) / (High - Low)

The average linguistic rating for each statement was calculated, leading to the creation of fuzzy numbers (L, M, H) representing the degree of consensus among the experts.

Fuzzy Number: The fuzzy number represents a range of agreement levels among the experts. It's provided in a triad format (lower bound, midpoint, upper bound), reflecting uncertainty or imprecision in their opinions.

Fuzzy Delphi Score (A) ≥ 0.5: The "Fuzzy Delphi Score" represents the degree of convergence or consensus among the experts' opinions for each technology. If this score is greater than or equal to 0.5, it signifies a level of consensus. Higher scores indicate stronger agreement.
Fuzzy Score (A) = (Sum of Numerical Values for each expert) / (Total number of experts)

The results presented varying levels of agreement among the experts for different statements.
Following the Fuzzy Delphi method, the Defuzzification process converted the fuzzy numbers (L, M, H) back into crisp numerical values using the center of gravity method:

Defuzzified Value (DV) = ((L × 0) + (M × 0.5) + (H × 1)) / (Sum of Fuzzy Numbers).

Items with a fuzzy score (A) equal to or higher than the specified alpha-cut value of 0.5 are considered eligible for retention.

(%) Expert Consensus ≥ 80%: This column indicates the percentage of experts who reached a consensus level of 80% or higher for each technology. A high percentage suggests a significant level of agreement among the experts.

Expert Consensus: This column summarizes whether there was a consensus among the experts for each technology. "Accept" indicates a consensus was reached, while "Reject" indicates a lack of consensus.

References

Abulibdeh, A., Zaidan, E., Abulibdeh, R.: Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: challenges, opportunities, and ethical dimensions. J. Clean. Prod. **437** (2024). https://doi.org/10.1016/j.jclepro.2023.140527

Ahmad, M.F., Wan Abdul Ghapar, W.R.G.: The era of artificial intelligence in Malaysian higher education: impact and challenges in tangible mixed-reality learning system toward self exploration education (SEE). Procedia Comput. Sci. **163**, 2–10 (2019). https://doi.org/10.1016/j.procs.2019.12.079

Aithal, P.S., Aithal, S.: Optimizing the use of artificial intelligence-powered GPTs as teaching and research assistants by professors in higher education institutions: a study on smart utilization. Int. J. Manag. Technol. Soc. Sci. (2023). https://doi.org/10.47992/ijmts.2581.6012.0326

AlDhaen, F.: The use of artificial intelligence in higher education - systematic review. In: COVID-19 Challenges to University Information Technology Governance (2022). https://doi.org/10.1007/978-3-031-13351-0_13

Ali, M., Abdel-Haq, M.K.: Bibliographical analysis of artificial intelligence learning in higher education: is the role of the human educator and educated a thing of the past? In: Fostering Communication and Learning With Underutilized Technologies in Higher Education (2020). https://doi.org/10.4018/978-1-7998-4846-2.ch003

Al-Maskari, A., Al Riyami, T., Ghnimi, S.: Factors affecting students' preparedness for the fourth industrial revolution in higher education institutions. J. Appl. Res. High. Educ. **16**(1) (2024). https://doi.org/10.1108/JARHE-05-2022-0169

Alnajim, A.: Transformation of education system in Saudi Arabia through artificial intelligence. SSRN Electron. J. (2023). https://doi.org/10.2139/ssrn.4362152

Altbach, P.G., Reisberg, L., Rumbley, L.E.: Trends in global higher education : tracking an academic revolution. In: UNESCO 2009 World Conference on Higher Education (2009). https://doi.org/10.1016/j.bse.2004.04.006

Androutsopoulou, A., Karacapilidis, N., Loukis, E., Charalabidis, Y.: Transforming the communication between citizens and government through AI-guided chatbots. Gov. Inf. Q. **36**(2) (2019). https://doi.org/10.1016/j.giq.2018.10.001

Antony, S., Ramnath, R.: A phenomenological exploration of students' perceptions of AI Chatbots in higher education. IAFOR J. Educ. **11**(2) (2023). https://doi.org/10.22492/ije.11.2.01

Ara Shaikh, A., Kumar, A., Jani, K., Mitra, S., García-Tadeo, D.A., Devarajan, A.: The role of machine learning and artificial intelligence for making a digital classroom and its sustainable impact on education during Covid-19. Mater. Today: Proc. **56**, 3211–3215 (2022). https://doi.org/10.1016/j.matpr.2021.09.368

Baradziej, S.: The role of AI algorithms in intelligent learning systems. In: AIET 2023. LNDE and CT, vol. 190 (2023). https://doi.org/10.1007/978-981-99-7947-9_14

Baskara, F.X.R.: Navigating pedagogical evolution: the implication of generative AI on the reinvention of teacher education. Prosiding Seminar Nasional Universitas Jabal ….Baskara, FX. R. (2023a). AI-Driven Dynamics: ChatGPT Transforming ELT Teacher-Student Interactions. Lensa: Kajian Kebahasaan, Kesusastraan, Dan Budaya, 13(2) (2023). https://doi.org/10.26714/lensa.13.2.2023.261-275

Berezinska, O.: Artificial intelligence in education: development and potential. In: Socio-humanitarian and Technical technological Explorations of Modern Science (2023). https://doi.org/10.51587/9798-9866-95983-2023-015-130-136

Bhutoria, A.: Personalized education and Artificial Intelligence in the United States, China, and India: a systematic review using a Human-In-The-Loop model. Comput. Educ. Artif. Intell. **3**, 100068 (2022). https://doi.org/10.1016/j.caeai.2022.100068

Bucea-Manea-ţoniş, R., et al.: Artificial intelligence potential in higher education institutions enhanced learning environment in Romania and Serbia. Sustainability (Switzerland) **14**(10) (2022). https://doi.org/10.3390/su14105842

Camaréna, S.: Artificial Intelligence (AI) For Sustainable Institutional Food Systems: Implementation of AI tools for school nutrition program management in the United States of America. Front. Sustain. Food Syst. **6**, 743810 (2022). https://doi.org/10.3389/fsufs.2022.743810

Caplan, B., et al.: How will artificial intelligence change higher ed? Chronicle High. Educ. **69**(20) (2023)

Cavus, N., Mohammed, Y.B., Yakubu, M.N.: Determinants of learning management systems during covid-19 pandemic for sustainable education. Sustainability (Switzerland) **13**(9) (2021a). https://doi.org/10.3390/su13095189

Creely, E., Blannin, J.: The implications of generative AI for creative composition in higher education and initial teacher education. ASCILITE Publications (2023). https://doi.org/10.14742/apubs.2023.618

Dautov, R., Husom, E.J., Sen, S., Song, H.: Towards community-driven generative AI. In: Position Papers of the 18th Conference on Computer Science and Intelligence Systems, vol. 36 (2023). https://doi.org/10.15439/2023f5494

Du, Z., Liang, X., Chen, S., Zhu, X., Chen, K., Jin, X.: Knowledge-infused deep learning diagnosis model with self-assessment for smart management in HVAC systems. Energy **263**, 125969 (2023). https://doi.org/10.1016/j.energy.2022.125969

El-Adl, A.M.: The potential of artificial intelligence to develop the education system in Egypt. المجلة, **34**(122) (2024a), المصرية للدراسات النفسية. https://doi.org/10.21608/ejcj.2024.336098

Eldeeb, A., Abdel Salam, S., Abdel Salam, M.R.: The foresight of economics of higher education in Egypt for achieving targets of sustainable development. J. Environ. Sci. **51**(2) (2022). https://doi.org/10.21608/jes.2022.108573.1129

Elimadi, I., Chafiq, N., Ghazouani, M.: Artificial intelligence in the context of digital learning environments (DLEs): towards adaptive learning (2024). https://doi.org/10.1007/978-3-031-50300-9_6

Et al., A.R.: A study on application of artificial intelligence and machine learning in Indian taxation system. Psychol. Educ. J. **58**(2) (2021). https://doi.org/10.17762/pae.v58i2.2265

Farazouli, A., Cerratto-Pargman, T., Bolander-Laksov, K., McGrath, C.: Hello GPT! Goodbye home examination? An exploratory study of AI Chatbots impact on university teachers' assessment practices. Assess. Eval. High. Educ. **49**(3) (2024). https://doi.org/10.1080/02602938.2023.2241676

Fernández-Gutiérrez, M., et al.: Self-management in heart failure using mHealth: a content validation. Int. J. Med. Informatics **171**, 104986 (2023). https://doi.org/10.1016/j.ijmedinf.2023.104986

Gašević, D., Siemens, G., Sadiq, S.: Empowering learners for the age of artificial intelligence. In: Computers and Education: Artificial Intelligence, vol. 4 (2023a). https://doi.org/10.1016/j.caeai.2023.100130

González-Calatayud, V., Prendes-Espinosa, P., Roig-Vila, R., Carpanzano, E.: Applied Sciences Review Artificial Intelligence for Student Assessment: A Systematic Review. (2021)

Hameed, H.A.: Artificial intelligence: what it was, and what it should be? Int. J. Adv. Comput. Sci. Appl. **11**(6) (2020). https://doi.org/10.14569/IJACSA.2020.0110609

Heilala, J., Shibani, A., de Freitas, A.G.: The requirements for heutagogical attunement within STEAM education. Int. J. Emerg. Technol. Learn. **18**(16) (2023). https://doi.org/10.3991/ijet.v18i16.42313

II, W.N.P.: Artificial intelligence in the medical system: four roles for potential transformation. Publ. Law Legal Theory Res. Paper Ser. **631** (2019)

Jeyaraman, M., Balaji, S., Jeyaraman, N., Yadav, S.: Unraveling the ethical enigma: artificial intelligence in healthcare. Cureus (2023). https://doi.org/10.7759/cureus.43262

Jones, J., Masika, R.: Appreciative inquiry as a developmental research approach for higher education pedagogy: space for the shadow. High. Educ. Res. Dev. **40**(2) (2021). https://doi.org/10.1080/07294360.2020.1750571

Kalyani, L.K.: Revolutionizing education: artificial intelligence's pioneering role in shaping tomorrow's scholars. Int. J. Multidisc. Res. Arts Sci. Technol. **1**(2) (2023). https://doi.org/10.61778/ijmrast.v1i2.6

Karimi, H., Khawaja, S.: The impact of artificial intelligence on higher education in England. Creat. Educ. **14**(12) (2023). https://doi.org/10.4236/ce.2023.1412154

La Torre, D., Appio, F.P., Masri, H., Lazzeri, F., Schiavone, F.: Impact of artificial intelligence in business and society. Impact Artif. Intell. Bus. Soc. (2023). https://doi.org/10.4324/9781003304616

Lainjo, B., Tmouche, H.: Meta-study of the evolutionary transformative academic landscape by artificial intelligence and machine learning. Int. J. Educ. Teach. Soc. Sci. **4**(1) (2024). https://doi.org/10.47747/ijets.v4i1.1626

Lakshmi, A.J., Kumar, A., Kumar, M.S., Patel, S.I., Naik, S.K.L., Ramesh, J.V.N.: Artificial intelligence in steering the digital transformation of collaborative technical education. J. High Technol. Manag. Res. **34**(2) (2023). https://doi.org/10.1016/j.hitech.2023.100467

Li, J., Zhou, Y., Yao, J., Liu, X.: An empirical investigation of trust in AI in a Chinese petrochemical enterprise based on institutional theory. Sci. Rep. **11**(1) (2021). https://doi.org/10.1038/s41598-021-92904-7

Mallik, S., Gangopadhyay, A.: Proactive and reactive engagement of artificial intelligence methods for education: a review. Front. Artif. Intell. **6**. (2023). https://doi.org/10.3389/frai.2023.1151391

Mcintosh, C.: Lifelong learning & distance higher education lifelong learning & distance higher education. Lifelong Learn. **4**(1) (2005)

Memarian, B., Doleck, T.: Fairness, accountability, transparency, and ethics (FATE) in artificial intelligence (AI) and higher education: a systematic review. Comput. Educ. Artif. Intell. **5** (2023). https://doi.org/10.1016/j.caeai.2023.100152

Mohsin, K.: Artificial intelligence - ethical principles & its positive impact on the society. Yapay Zekâ - Etik İlkeler & Toplumdaki Olumlu Etkileri., **8**(16) (2020)

Naidu, A.: Impact of Artificial Intelligence on Society. Indian Institute of Science (2019)

Nyanga, P.H.: Factors influencing adoption and area under conservation agriculture: a mixed methods approach. Sustain. Agric. Res. **1**(2) (2012). https://doi.org/10.5539/sar.v1n2p27

Paton, R.: Digital evolution: a new approach to learning and teaching in higher education | Times Higher Education (THE). Times Higher Education (2018)

Peterson, M., Keller, G.: Academic strategy: the management revolution in higher education. J. High. Educ. **55**(5) (1984). https://doi.org/10.2307/1981829

Photopoulos, P., Triantis, D.: Think twice: first for tech, then for ed. SN Comput. Sci. **4**(2) (2023a). https://doi.org/10.1007/s42979-022-01538-7

Piddubtseva, O.: The potential of use of artificial intelligence technologies in institutions of higher education. Human. Sci. Curr. Issues **3**(71) (2024). https://doi.org/10.24919/2308-4863/71-3-33

Pillai, R., Sivathanu, B., Metri, B., Kaushik, N.: Students' adoption of AI-based teacher-bots (T-bots) for learning in higher education. Inf. Technol. People **37**(1) (2024). https://doi.org/10.1108/ITP-02-2021-0152

Qian, C.: Research on human-centered design in college music education to improve student experience of artificial intelligence-based information systems. J. Inf. Syst. Eng. Manag. **8**(3) (2023). https://doi.org/10.55267/iadt.07.13854

Ranga, M.: Artificial intelligence and its impact on society. Int. J. Eng. Sci. Human. **14**(Special Issue 1) (2024). https://doi.org/10.62904/9fmpr623

Rudolph, J., Tan, S., Tan, S.: War of the Chatbots: Bard, Bing Chat, ChatGPT, Ernie and beyond. The new AI gold rush and its impact on higher education. J. Appl. Learn. Teach. **6**(1) (2023b). https://doi.org/10.37074/jalt.2023.6.1.23

Sabharwal, D., Kabha, R., Srivastava, K.: Artificial Intelligence (AI)-powered virtual assistants and their effect on human productivity and laziness: study on students of Delhi-NCR (India) & Fujairah (UAE). J. Content Community Commun. **17**(9) (2023). https://doi.org/10.31620/JCCC.06.23/12

Sandu, N., Gide, E.: Adoption of AI-chatbots to enhance student learning experience in higher education in India. In: 2019 18th International Conference on Information Technology Based Higher Education and Training, ITHET 2019 (2019a). https://doi.org/10.1109/ITHET46829.2019.8937382

Sarwar, M.A., Saima, M., Gul, A.: The role of artificial intelligence in shaping the future of education at higher secondary level. J. Educ. Soc. Stud. **5**(1) (2024). https://doi.org/10.52223/jess.2024.5104

Shamkuwar, M., Jindal, P., More, R., Patil, P., Mahamuni, P.: Artificial intelligence and higher education: a systematic visualizations based review. J. E-Learn. Knowl. Soc. **19**(3) (2023). https://doi.org/10.20368/1971-8829/1135857

Sharma, M., Luthra, S., Joshi, S., Kumar, A.: Implementing challenges of artificial intelligence: evidence from public manufacturing sector of an emerging economy. Gov. Inf. Q. **39**(4) (2022a). https://doi.org/10.1016/j.giq.2021.101624

Siddique, S., Chow, J.C.L.: Machine learning in healthcare communication. Encyclopedia **1**(1) (2021). https://doi.org/10.3390/encyclopedia1010021

Singh, S.V., Hiran, K.K.: The impact of AI on teaching and learning in higher education technology. J. High. Educ. Theory Pract. **22**(13) (2022). https://doi.org/10.33423/jhetp.v22i13.5514

Slowey, M., Kozina, E.: Perspectives on professional development: the voice of Irish academics. In: Forming, Recruiting and Managing the Academic Profession (2015). https://doi.org/10.1007/978-3-319-16080-1_15

SN, S.: Empowering learners in the digital age: Integrating technology and personalized learning in higher education. Int. J. Adv. Acad. Stud. **5**(10) (2023). https://doi.org/10.33545/27068919.2023.v5.i10a.1059

Sok, S., Heng, K.: Opportunities, challenges, and strategies for using ChatGPT in higher education: a literature review. J. Digit. Educ. Technol. **4**(1) (2023). https://doi.org/10.30935/jdet/14027

Useem, A.: "Higher-education revolution" divides academics in Sudan. Chronicle High. Educ. **44**(25) (1998)

Vargas-Murillo, A.R., Pari-Bedoya, I.N.M.D.L.A., Guevara-Soto, F.D.J.: The ethics of AI assisted learning: a systematic literature review on the impacts of ChatGPT usage in education. In: ACM International Conference Proceeding Series (2023). https://doi.org/10.1145/3606094.3606101

Vial, G., Cameron, A.F., Giannelia, T., Jiang, J.: Managing artificial intelligence projects: key insights from an AI consulting firm. Inf. Syst. J. **33**(3) (2023). https://doi.org/10.1111/isj.12420

Wang, T.: Navigating generative AI (ChatGPT) in higher education: opportunities and challenges. In: LNET, Part F2953 (2023). https://doi.org/10.1007/978-981-99-5961-7_28

Yadav, N., Sharma, L., Dhake, U.: Artificial intelligence: the future. Int. J. Sci. Res. Eng. Manag. **07**(12) (2023a). https://doi.org/10.55041/ijsrem27796

Yang, S., Evans, C.: Opportunities and challenges in using AI chatbots in higher education. In: ACM International Conference Proceeding Series (2019a). https://doi.org/10.1145/3371647.3371659

Yu, P., Mulli, J., Syed, Z.A.S., Umme, L.: Facilitating global collaboration and knowledge sharing in higher education with generative AI. In: Facilitating Global Collaboration and Knowledge Sharing in Higher Education With Generative AI (2023a). https://doi.org/10.4018/9798369304877

Zhang, H., Lee, I., Ali, S., DiPaola, D., Cheng, Y., Breazeal, C.: Integrating ethics and career futures with technical learning to promote AI literacy for middle school students: an exploratory study. Int. J. Artif. Intell. Educ. **33**(2) (2023a). https://doi.org/10.1007/s40593-022-00293-3

Zmyzgova, T.R., Polyakova, E.N., Karpov, E.K.: Digit. Transform. Educ. Artif. Intell. (2020). https://doi.org/10.2991/aebmr.k.200502.134

IoT Integration for Smart Agriculture: Prioritizing Key Factors Using WASPAS Approach

M. K. P. Naik[1], Anand Jaiswal[2(✉)], Ashish Yadav[1], and Teena Singh[1]

[1] New Delhi Institute of Management, New Delhi 110062, India
[2] TERI School of Advanced Studies, New Delhi 110070, India
anand.jaiswal27@gmail.com

Abstract. The study aims to investigate IoT integration for smart agriculture by identifying and ranking significant factors impacting IoT adoption. It emphasizes the importance of prioritizing factors in order to ensure optimal implementation. The study uses the MCDM approach of the Weighted Aggregated Sum Product Assessment (WASPAS) to comprehensively analyze thirteen key factors, crucial for the efficient integration of IoT in smart agriculture, identified through literature review. Data was collected on a 5-point ranking scale from farmers, government experts, research, and academicians. The findings suggest that skill shortages and high capital investment costs are significant barriers to IoT implementation in this context. Addressing these barriers is critical to promoting widespread adoption, emphasizing the need to close skill shortages and investigate cost-effective alternatives along with other identified factors. The study's findings offer practical recommendations for stakeholders, policymakers, and agricultural practitioners, enabling them to overcome barriers and accelerate the integration of IoT technologies into agricultural practices, ultimately improving efficiency and sustainability in food production.

Keywords: IoT · Smart Agriculture · WASPAS

1 Introduction

Smart agriculture, also known as digital farming or precision agriculture, has become increasingly important in today's world due to its ability to address various challenges in agricultural production. It focuses on enhancing productivity, efficiency, and sustainability in farming practices (Wang et al., 2021). By integrating advanced technologies such as Internet of Things (IoT), artificial intelligence (AI), and machine learning (ML), smart agriculture offers solutions for optimizing agricultural processes, reducing waste, and increasing yields (Gupta et al., 2023). It plays a crucial role in promoting sustainable agricultural development by enabling farmers to make informed decisions based on real-time data and analytics (Wang et al., 2021). Through the use of technologies like IoT-based systems for farm management, smart agriculture enhances the efficiency and intelligence of agricultural production (Gao et al., 2023). It also facilitates the adoption

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 452–463, 2025.
https://doi.org/10.1007/978-3-031-86305-9_31

of responsible research and innovation practices in farming, addressing socio-ethical challenges associated with technological advancements in agriculture (Eastwood et al., 2017).

The adoption and usability of smart agriculture practices in India also have been gaining traction in recent years. Various studies have highlighted the significance of smart agriculture in enhancing agricultural sustainability and productivity in the country (Aryal et al., 2018). Climate change poses a significant threat to Indian agriculture, making the adoption of climate-smart agricultural practices crucial for ensuring food security and mitigating the impacts of climate variability (Aryal et al., 2018; Kishore et al., 2018). Studies have also emphasized the role of ICT (Information and Communication Technology) in optimizing resource usage and increasing crop yields through smart agriculture practices in India (Bouali et al., 2022; Yang et al., 2021). The integration of IoT technologies in agriculture monitoring and control systems has the potential to revolutionize farming practices and improve efficiency (Yang et al., 2021; Pawar et al., 2022). Also, the use of smartphone applications for agro-advisory services is becoming increasingly popular among farmers in India, enabling them to make informed decisions on climate-smart natural resource management (Singh et al., 2020).

Despite the potential benefits of smart agriculture, the adoption level in India faces challenges related to technological constraints, economic factors, and other barriers (Aggarwal et al., 2021). However, the government of India has shown commitment by investing a substantial amount towards the development of smart agriculture in the country (Kishore et al., 2018). Still, there is a substantial need to examine the factors and causes preventing smart agriculture adoption in India. With particular relevance to IoT, India has made the internet affordable and readily accessible. Most rural areas now have connectivity, and farmers are also using smartphones and IoT devices. Consequently, there is a strong need to assess and prioritize the factors critical for the proper implementation of IoT integration in smart agriculture in India, to enhance productivity and sustainability.

The significance of this study lies in its contribution to addressing the issues faced by agriculture in India through the perspective of smart agriculture and IoT integration. Adopting smart agriculture practices becomes crucial as the country with issues related to climate change, resource scarcity, and the need for increased productivity. By prioritizing key factors using the WASPAS (Weighted Aggregated Sum Product Assessment) approach, this study aims to provide actionable insights for policymakers, stakeholders, and agricultural practitioners. Through this study, we hope to contribute to the advancement of sustainable agricultural development in India, ultimately ensuring food security and mitigating the impacts of operational challenges and climate change on the agricultural sector.

2 Literature Review and Research Framework

Using IoT technologies to allow smart agriculture presents a possible way to advance agricultural operations. However, a number of barriers prevent IoT from being seamlessly incorporated into smart agriculture. Research by Strong et al. (2022) examines

the adoption barriers for IoT smart agriculture technologies, emphasizing the importance of understanding stakeholder characteristics and communication barriers. Similarly, Padyab et al. (2019) investigate IoT adoption barriers across various domains, including smart agriculture, exploring on challenges faced in large-scale pilots. In context to this, Shrestha (2024) provides insights into the roles of real-time nitrogen sensing and IoT integration in smart agriculture, highlighting both opportunities and challenges associated with these technologies. Pranto et al. (2021) focus on the integration of blockchain and smart contracts with IoT devices in agriculture, emphasizing the potential benefits in pre-harvesting and post-harvesting segments. Additionally, Gupta et al. (2022) discusses the applications of IoT and machine learning in smart agriculture, identifying research gaps and advancements in this field. These studies collectively underscore the significance of overcoming barriers to IoT integration in smart agriculture to realize its full potential in revolutionizing agricultural practices. However, there is also a need to assess and prioritize the factors (barriers) that hinder IoT integration for smart agriculture in the Indian context.

2.1 Factors of IoT Integration for Smart Agriculture

Using IoT technologies to allow smart agriculture presents a possible way to advance agricultural operations. However, a number of barriers as listed in Table 1 prevent IoT from being seamlessly incorporated into smart agriculture. Research by Strong et al. (2022) examines the adoption barriers for IoT smart agriculture technologies, emphasizing the importance of understanding stakeholder characteristics and communication barriers. Similarly, Padyab et al. (2019) investigate IoT adoption barriers across various domains, including smart agriculture, exploring on challenges faced in large-scale pilots. In context to this, Shrestha (2024) provides insights into the roles of real-time nitrogen sensing and IoT integration in smart agriculture, highlighting both opportunities and challenges associated with these technologies. Pranto et al. (2021) focus on the integration of blockchain and smart contracts with IoT devices in agriculture, emphasizing the potential benefits in pre-harvesting and post-harvesting segments. Additionally, Gupta et al. (2022) discusses the applications of IoT and machine learning in smart agriculture, identifying research gaps and advancements in this field. These studies collectively underscore the significance of overcoming barriers to IoT integration in smart agriculture to realize its full potential in revolutionizing agricultural practices. However, there is also a need to assess and prioritize the factors (barriers) that hinder IoT integration for smart agriculture in the Indian context. To investigate it, the study follows the research framework to prioritize the factors of IoT integration for smart agriculture, as provided in Fig. 1.

3 Methodology and Data Collection

The methodology employed for this study involved the utilization of the WASPAS approach to prioritize key factors. It is a multi-criteria decision-making (MCDM) method used to evaluate alternatives based on multiple criteria. It combines the advantages of both the Weighted Sum Model (WSM) and the Weighted Product Model (WPM) to provide a comprehensive decision-making framework.

Table 1. Factors identified by literature review

Factors	Definition	Refence
F1. Lack of standardized standards	Overcoming the absence of uniform guidelines and protocols for IoT implementation, hindering interoperability and collaboration among diverse systems	Waqar et al., 2023; Narwane et al., 2022
F2. Security and privacy	Addressing apprehensions surrounding the protection of sensitive agricultural data and ensuring compliance with stringent privacy regulations	Quy et al., 2022; Adli et al., 2023;
F3. Reliability	Ensuring consistent and dependable performance of IoT systems in varying agricultural environments to maintain operational efficiency	Farooq et al., 2019; Morchid et al., 2023;
F4. Lack of skills/ knowledge	Bridging the gap in technical expertise and understanding among farmers and agricultural stakeholders for effective utilization of IoT technologies	Narwane et al., 2022; Waqar et al., 2023
F5. Coverage and connectivity	Tackling challenges related to inadequate internet connectivity and network coverage in rural agricultural areas, essential for seamless data transmission and communication	Ayaz et al., 2019; Narwane et al., 2022
F6. Legal and regulatory requirements	Navigating complex legal and regulatory landscapes to ensure adherence to data protection laws and agricultural policies	Waqar et al., 2023; Elijah et al., 2018;
F7. High capital investment cost	Managing the substantial upfront expenses associated with procuring and implementing IoT infrastructure and solutions	Dhanaraju et al., 2022; Morchid et al., 2023
F8. Fear of new technology	Addressing apprehension and reluctance among farmers and agricultural stakeholders toward adopting unfamiliar IoT technologies, often due to concerns about usability, reliability, and potential disruptions to existing practices	Narwane et al., 2022; Ayaz et al., 2019

(*continued*)

In this study, the formulation of the WASPAS method involved several steps. Firstly, criteria values were normalized to ensure comparability across different criteria. This was

Table 1. (*continued*)

Factors	Definition	Refence
F9. Operational cost	Addressing ongoing expenditures, including maintenance, energy consumption, and support services required for the sustainable operation of IoT systems	Quy et al., 2022
F10. Inadequate government policies and incentives	Advocating for supportive policy frameworks and incentives to encourage widespread adoption of IoT in agriculture and stimulate innovation	Waqar et al., 2023
F11. Resource Optimization	Farmers face complexity in optimizing resource allocation for IoT devices, cloud storage, and data transmission due to varying farm sizes and monitoring needs. This requires advanced mathematical models and algorithms to maximize agriculture productivity efficiently	Farooq et al., 2019; Elijah et al., 2018
F12. Interference	Mitigating interference problems arising from the proliferation of IoT devices and wireless communication technologies in agricultural environments	Quy et al., 2022; Dhanaraju et al., 2022
F13. Choice of technology	Navigating the complexity of selecting the most suitable IoT solutions amidst a plethora of available options, considering factors such as compatibility, scalability, and long-term viability	Elijah et al., 2018; Ayaz et al., 2019

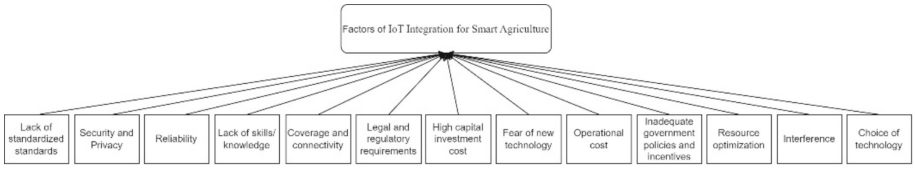


Fig. 1. Research framework to prioritize the factors of IoT integration for smart agriculture

achieved by transforming the data collected from the experts on a common scale, ranging from 0 to 1. Next, weights were assigned to each criterion to reflect its importance in the assessment process. These weights were determined based on the expertise and relevance of experts. In our study, recognizing the significance of farmers and government officials, a weight of 0.4 was assigned to each of them, while a weight of 0.2 was assigned to academicians and researchers. Subsequently, the aggregated performance score for each alternative was calculated by multiplying the normalized values of each criterion by

its corresponding weight and summing the products. This step integrated the weighted scores to represent the overall performance of each alternative. Finally, alternatives were ranked based on their aggregated scores, with the alternative having the highest aggregated score considered the most favorable or optimal solution.

3.1 Sample and Data Collection

The data collection process involved judgmental sampling, where stakeholders including farmers, government officials, and researchers/academicians specializing in agriculture were selectively chosen based on their expertise and relevance to the study. For the study, seven farmers, four government experts, and four academicians/researchers were consulted. The selection criteria ensured expertise and representation from diverse perspectives within the agriculture sector. Prior to participation, all stakeholders were provided with detailed information about the study objectives and procedures, and informed consent was obtained from each participant. The data collection phase took place from January 2024 to March 2024. During consultations, participants were asked to rate the priority of each identified factor on a 5-point rating scale, enabling an assessment of their significance in the context of IoT integration in smart agriculture in India.

4 Findings and Discussion

This section discusses the prioritization of key factors influencing IoT integration in smart agriculture based on the data collected from stakeholders and the application of WASPAS method. The findings reveal insights into the perceptions and priorities of farmers, government officials, and researchers/academicians within the agriculture sector. Through the analysis of rankings provided by different stakeholders, as well as the normalization, weight assignment, and aggregation processes, an understanding of the factors hindering IoT integration emerges.

Table 2 presents the ratings assigned by different experts to various factors influencing IoT integration in smart agriculture. Ratings were provided by farmers, government officials, and researchers/academicians. Notably, factors such as Lack of skills/Knowledge, High capital investment cost, and Operational cost received high ratings across all expert groups, indicating their significant impact on IoT integration in smart agriculture.

Following the ratings provided by different experts, a Normalization Matrix was developed to ensure comparability across the various factors. Table 3 presents the Normalization matrix of ratings, where each factor's ratings provided by farmers, government officials, and researchers/academicians were normalized. Normalization allows for the transformation of raw ratings into a common scale, facilitating a standardized comparison of the factors' importance across different expert groups. The values in the matrix represent the normalized ratings for each factor, with scores ranging between 0 and 1.

After the development of the Normalization Matrix, the Weighted Sum Model (WSM) was applied to calculate the aggregated scores for each factor based on their normalized ratings and the assigned weights. Table 4 presents the results of the Weighted

Table 2. Ratings assigned by different experts to identified factors

Factors	Ratings by different expects		
	Farmer	*Government*	*Researcher/Academicians*
Lack of standardized standards (F1)	3.00	2.75	3.00
Security and Privacy (F2)	2.86	2.50	2.75
Reliability (F3)	4.14	3.50	3.75
Lack of skills/Knowledge (F4)	4.71	4.50	4.50
Coverage and connectivity (F5)	4.00	3.25	3.50
Legal and regulatory requirements (F6)	3.14	2.75	3.00
High capital investment cost (F7)	4.57	4.25	4.25
Fear of new technology (F8)	3.86	3.00	3.25
Operational cost (F9)	4.43	4.00	4.25
Inadequate government policies and incentives (F10)	3.29	3.00	3.25
Resource Optimization (F11)	4.29	3.75	4.00
Interference (F12)	3.71	3.00	3.25
Choice of technology (F13)	3.57	3.00	3.25

Sum Model, (WSM) for farmers, government officials, and researchers/academicians. In this matrix, the normalized ratings for each factor are multiplied by their corresponding weights and summed, providing an aggregated score that reflects the overall importance of each factor in influencing IoT integration in smart agriculture according to different expert groups.

Following the Weighted Sum Model, the Weighted Product Model (WPM) was employed to calculate the aggregated scores for each factor based on their normalized ratings and the assigned weights. Table 5 presents the results of the Weighted Product Model, for farmers, government officials, and researchers/academicians. In this matrix, the normalized ratings for each factor are multiplied by their corresponding weights, providing an aggregated score that reflects the overall importance of each factor in influencing IoT integration in smart agriculture according to different expert groups.

The findings from the Preference Score and WASPAS Coefficient analysis, considering lambda as 0.5, are summarized in Table 6. The Preference Score of the WSM and WPM reflects the overall preference for each model, indicating their effectiveness in evaluating the factors influencing IoT integration in smart agriculture. Both models demonstrate comparable preference scores, suggesting that they offer similar insights into the prioritization of factors. Also, the WASPAS Coefficient represents the coefficient value obtained through the WASPAS approach, which combines the results from both models. The WASPAS Coefficient serves as a comprehensive measure of the overall importance of factors, considering the preferences derived from both the WSM and

Table 3. Normalization matrix of ratings

Factors	Normalization matrix		
	Farmer	*Government*	*Researcher/Academicians*
F1	0.64	0.61	0.67
F2	0.61	0.56	0.61
F3	0.88	0.78	0.83
F4	1.00	1.00	1.00
F5	0.85	0.72	0.78
F6	0.67	0.61	0.67
F7	0.97	0.94	0.94
F8	0.82	0.67	0.72
F9	0.94	0.89	0.94
F10	0.70	0.67	0.72
F11	0.91	0.83	0.89
F12	0.79	0.67	0.72
F13	0.76	0.67	0.72

Table 4. Results of the Weighted Sum Model

Factors	Normalized weighted matrix-WSM		
	Farmer	*Government*	*Researcher/Academicians*
F1	0.25	0.24	0.13
F2	0.24	0.22	0.12
F3	0.35	0.31	0.17
F4	0.40	0.40	0.20
F5	0.34	0.29	0.16
F6	0.27	0.24	0.13
F7	0.39	0.38	0.19
F8	0.33	0.27	0.14
F9	0.38	0.36	0.19
F10	0.28	0.27	0.14
F11	0.36	0.33	0.18
F12	0.32	0.27	0.14
F13	0.30	0.27	0.14

Table 5. Results of the Weighted Product Model

Factors	Normalized weighted matrix-WPM		
	Farmer	*Government*	*Researcher/Academicians*
F1	0.83	0.82	0.92
F2	0.82	0.79	0.91
F3	0.95	0.90	0.96
F4	1.00	1.00	1.00
F5	0.94	0.88	0.95
F6	0.85	0.82	0.92
F7	0.99	0.98	0.99
F8	0.92	0.85	0.94
F9	0.98	0.95	0.99
F10	0.87	0.85	0.94
F11	0.96	0.93	0.98
F12	0.91	0.85	0.94
F13	0.89	0.85	0.94

WPM. In this analysis, the WASPAS Coefficient consistently aligns with the preference scores obtained from both models, further validating its effectiveness in assessing the significance of factors in IoT integration decisions.

Based on the preference score and WASPAS coefficients obtained from the analysis, ranks were assigned to each factor, as shown in Table 6. The factors were ranked based on their aggregated scores, reflecting their perceived importance in influencing IoT integration in smart agriculture. The findings reveal that 'Lack of skills/Knowledge' emerged as the top-ranked factor, underscoring the critical importance of addressing skill gaps and enhancing technical expertise among farmers and agricultural stakeholders. Following closely behind, 'High capital investment cost' secured the second position, highlighting the substantial upfront expenses associated with procuring and implementing IoT infrastructure and solutions. 'Operational cost' secured the third position, emphasizing the ongoing expenditures required for the sustainable operation of IoT systems. Interestingly, factors such as 'Security and Privacy' and 'Legal and regulatory requirements' obtained relatively lower ranks, indicating that while they are significant, they may be perceived as less critical compared to other factors in the context of IoT integration in smart agriculture.

4.1 Implications of the Study

The prioritization of key factors in IoT integration for smart agriculture carries significant implications across various domains. From theoretical perspective, this study contributes to the existing body of knowledge by highlighting critical barriers and challenges in the

Table 6. Preference score, WASPAS coefficient and prioritization rank of factors

Preference score of WSM	Preference score of WPM	lambda	WASPAS Coefficient	Rank
0.63	0.63	0.5	0.63	12
0.59	0.59	0.5	0.59	13
0.83	0.83	0.5	0.83	5
1.00	1.00	0.5	1.00	1
0.78	0.78	0.5	0.78	6
0.64	0.64	0.5	0.64	11
0.95	0.95	0.5	0.95	2
0.74	0.74	0.5	0.74	7
0.92	0.92	0.5	0.92	3
0.69	0.69	0.5	0.69	10
0.87	0.87	0.5	0.87	4
0.73	0.72	0.5	0.73	8
0.71	0.71	0.5	0.71	9

adoption of IoT technology for smart agriculture in Indian context. The findings serve as a reference point for researchers and scholars interested in understanding the dynamics of technology adoption and its impact on agricultural practices (Aryal et al., 2018; Ayaz et al., 2019; Gupta et al., 2022; Quy et al., 2022).

From a practical standpoint, the implications extend to policymakers, agricultural practitioners, and technology developers. By identifying factors such as the lack of skills/knowledge and high capital investment costs (Narwane et al., 2022; Dhanaraju et al., 2022) as primary barriers, stakeholders can devise targeted interventions and policy measures. For example, government organization can allocate resources towards training programs to enhance digital literacy among farmers (Adli et al., 2023), while corporate sector can explore innovative financing models to mitigate the financial burden associated with technology adoption (Elijah et al., 2018). The insights derived from this study also offer actionable guidance for technology developers in refining IoT solutions tailored to the needs of the agricultural sector (Farooq et al., 2019). By addressing concerns related to training, security, privacy, and reliability, developers can improve greater trust and acceptance of IoT technologies among rural farmers for incorporating smart agriculture (Quy et al., 2022; Morchid et al., 2023).

4.2 Limitations and Future Scope

While this study provides valuable insights into the prioritization of key factors for IoT integration in smart agriculture, some limitations also exist. The study's scope was confined to Indian geographical context, potentially limiting the generalizability of the findings to other regions or countries with different socio-economic conditions and agricultural landscapes. Also, the data collection process relied on a judgmental sampling

approach, which may cause biases and limit the representativeness of the sample. In addition to it, the study focused on assessing the perceived importance of factors without investigating into the possible reasons of these perceptions from experts. Future research could address these limitations by employing larger and more diverse samples, applying random sampling techniques, and conducting qualitative explorations to gain deeper insights into stakeholders' perspectives and experiences. Also, exploring the intersectionality of factors such as farmer's financial status, education level, demography and other factors could provide a more practical understanding of the differential barriers and opportunities for IoT adoption among various stakeholder groups in agriculture.

References

Adli, H.K., et al.: Recent advancements and challenges of AIoT application in smart agriculture: a review. Sensors **23**(7), 3752 (2023)

Aggarwal, A., Gupta, S., Jamwal, A., Agrawal, R., Sharma, M., Dangayach, G.: Adoption of smart and sustainable manufacturing practices: an exploratory study of Indian manufacturing companies. Proc. Inst. Mech. Eng. Part B J. Eng. Manuf. **236**(5), 586–602 (2021)

Aryal, J.P., et al.: Adoption of multiple climate-smart agricultural practices in the gangetic plains of Bihar, India. Int. J. Clim. Change Strateg. Manag. **10**(3), 407–427 (2018)

Ayaz, M., Ammad-Uddin, M., Sharif, Z., Mansour, A., Aggoune, E.H.M.: Internet-of-Things (IoT)-based smart agriculture: toward making the fields talk. IEEE Access **7**, 129551–129583 (2019)

Bouali, E., Abid, M., Boufounas, E., Hamed, T., Benhaddou, D.: Renewable energy integration into cloud & IoT-based smart agriculture. IEEE Access **10**, 1175–1191 (2022)

Dhanaraju, M., Chenniappan, P., Ramalingam, K., Pazhanivelan, S., Kaliaperumal, R.: Smart farming: Internet of Things (IoT)-based sustainable agriculture. Agriculture **12**(10), 1745 (2022)

Eastwood, C., Klerkx, L., Ayre, M., Rue, B.: Managing socio-ethical challenges in the development of smart farming: from a fragmented to a comprehensive approach for responsible research and innovation. J. Agric. Environ. Ethics **32**(5–6), 741–768 (2017)

Elijah, O., Rahman, T.A., Orikumhi, I., Leow, C.Y., Hindia, M.N.: An overview of Internet of Things (IoT) and data analytics in agriculture: benefits and challenges. IEEE Internet Things J. **5**(5), 3758–3773 (2018)

Farooq, M.S., Riaz, S., Abid, A., Abid, K., Naeem, M.A.: A survey on the role of IoT in agriculture for the implementation of smart farming. IEEE Access **7**, 156237–156271 (2019)

Gao, J., Wu, R., Hao, J., Xu, C., Guo, H., Wang, H.: Energy-efficient resource scheduling and computation offloading strategy for solar-powered agriculture WSN. J. Sens. **2023**, 1–17 (2023)

Gupta, G., Abrol, V., Pradhan, S.: Smart farming: boosting crop management with SVM and random forest (2023)

Gupta, L., Malhotra, S., Kumar, A.: Study of applications of Internet of Things and Machine Learning for Smart Agriculture. In: 2022 IEEE International Conference on Current Development in Engineering and Technology (CCET), pp. 1–5. IEEE (2022)

Jarial, S.: Internet of Things application in Indian agriculture, challenges and effect on the extension advisory services–a review. J. Agribus. Dev. Emerg. Econ. (ahead-of-print) (2022)

Kishore, A., Pal, B., Joshi, K., Aggarwal, P.: Unfolding government policies towards the development of climate smart agriculture in India. Agric. Econ. Res. Rev. **31**(conf), 123 (2018)

Kumar, S., Tiwari, P., Zymbler, M.: Internet of Things is a revolutionary approach for future technology enhancement: a review. J. Big Data **6**(1), 1–21 (2019)

Morchid, A., El Alami, R., Raezah, A.A., Sabbar, Y.: Applications of internet of things (IoT) and sensors technology to increase food security and agricultural Sustainability: benefits and challenges. Ain Shams Eng. J. 102509 (2023)

Narwane, V.S., Gunasekaran, A., Gardas, B.B.: Unlocking adoption challenges of IoT in Indian agricultural and food supply chain. Smart Agric. Technol. **2**, 100035 (2022)

Padyab, A., Habibipour, A., Rizk, A., Ståhlbröst, A.: Adoption barriers of IoT in large scale pilots. Information **11**(1), 23 (2019)

Pawar, M., More, V., Jalke, P., Nichal, A.: IoT based smart agriculture monitoring, weather control and irrigation system. Int. J. Res. Public. Rev. **03**(12), 1688–1694 (2022)

Pranto, T., Noman, A., Mahmud, A., Haque, A.: Blockchain and smart contract for IoT enabled smart agriculture. PeerJ Comput. Sci. **7**, e407 (2021)

Quy, V.K., et al.: IoT-enabled smart agriculture: architecture, applications, and challenges. Appl. Sci. **12**(7), 3396 (2022)

Shrestha, M.: Review—perspectives on the roles of real time nitrogen sensing and IoT integration in smart agriculture. J. Electrochem. Soc. **171**(2), 027526 (2024)

Singh, R., Singh, R., Singh, N., Anurag, T., Koyu, B.: Application of smart phone agro-advisory services of m4agrinei in climate smart natural resource management in agriculture by tribal farmers of Meghalaya: an empirical study with structural equation modeling. J. Agric. Ecol. **09**(01), 67–77 (2020)

Strong, R., Wynn, J., Lindner, J., Palmer, K.: Evaluating Brazilian agriculturalists' IoT smart agriculture adoption barriers: understanding stakeholder salience prior to launching an innovation. Sensors **22**(18), 6833 (2022)

Tzounis, A., Katsoulas, N., Bartzanas, T., Kittas, C.: Internet of Things in agriculture, recent advances and future challenges. Biosys. Eng. **164**, 31–48 (2017)

Wang, T., et al.: Applications of UAS in crop biomass monitoring: a review. Front. Plant Sci. 12 (2021)

Waqar, A., Khan, M.B., Shafiq, N., Skrzypkowski, K., Zagórski, K., Zagórska, A.: Assessment of challenges to the adoption of IoT for the safety management of small construction projects in Malaysia: structural equation modeling approach. Appl. Sci. **13**(5), 3340 (2023)

Yang, J., Sharma, A., Kumar, R.: Iot-based framework for smart agriculture. Int. J. Agric. Environ. Inf. Syst. **12**(2), 1–14 (2021)

Transforming EHR Systems Using the HyperLedger Fabric Blockchain Framework

Neha A. Samsir(✉) and Arpit A. Jain

GLS University, Ahmedabad, Gujarat, India
{neha.samsir,arpit.jain}@glsuniversity.ac.in

Abstract. An Internet-connected network of nodes running a distributed protocol called a blockchain simulates a "trusted" computing service. Each node owns a portion of the asset that the service generates or represents. Blockchain technology's main objective is the decentralization of information and data management and storage. The popularity of permissioned blockchain platforms has significantly increased recently. One such permissioned blockchain technology is called Hyperledger Fabric, and It runs as one of the HyperLedger projects under the Linux Desktop Federation's administration. A modular and extendable open-source method for launching projects is called Fabric. Smart contracts (chaincode), endorsers, committers, validators, andOrderers are among the constituents comprising the Fabric. These platforms maintain a high degree of anonymity for their users while guaranteeing total transparency of the activities documented on the ledger, data security, and resistance to tampering.

Keywords: blockchain · hyperledger fabric · chaincode · EHR · privacy · decentralized file system

1 Introduction

Blockchain technology has the potential to significantly improve the effectiveness, privacy, and security of Electronic Health Records (EHR). Existing centralized database-based solutions are vulnerable to well-known security issues like Denial of Service (DoS) attacks and single points of failure that resemble conventional database systems. An immutable ledger for transaction recording is created by a distributed, mutually distrusting network of peers called a distributed ledger system called blockchain. A copy of the ledger is kept on file or transaction by each peer. A consensus procedure must be followed by each peer to be able arrange the transactions into blocks, which must be created into a hash chain over the blocks. The transaction is shown as a collection of blocks, and each node has an exact duplicate of it. Each block in the ledger contains the preceding one's hash, ensuring its immutability. Blockchains can function as permissioned or permissionless networks. The consensus protocol is used to ensure the consistency of all nodes. The two primary types of blockchain management are private as well as public blockchains. Blockchain technology has the potential to significantly improve Electronic Health Record (EHR) sharing systems' security, privacy, and efficiency.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 464–476, 2025.
https://doi.org/10.1007/978-3-031-86305-9_32

HyperLedger Fabric is an open-sourced permissioned blockchain. It is intended for utilization in applications at the organizational level. A high level of security, privacy, and scalability is offered by the platform. It facilitates the development of specialized blockchain solutions for a range of use cases across numerous industries, including finance, healthcare, and supply chain. Each node in the HyperLedger Fabric network has a specific task to accomplish, such as validating transactions, maintaining the ledger, or processing chaincode. A basic explanation of the HyperLedger Fabric is given below. Patients' electronic health records, or EHRs, are digitally preserved medical records. The health record contains primarily administrative data. It includes personal information about the patient, for example name, address, age, sex, phone number, and so on moreover, clinical data, such as the patient's lab results, diagnosis summary, and other information. One of the numerous advantages of switching to electronic health records from paper ones is the lower possibility of theft or loss. To maintain the requisite privacy in the suggested system, Hyperledger Fabric is utilized. Since patient is the network owner and has full access rights, This network has a high level of permissions. Therefore, medical staff will possess the ability to access a portion of the patient's digital health record in an emergency thanks to the backup access mechanism built into the proposed system.

A system of electronic health records or EHRS, is an electronic version of keeping a patient's medical record online, containing extremely private and secret details regarding the patient's medical history, diagnosis, and treatment plan. Appointment records, Prescriptions, and lab test results are typical instances of EHR data that should be retained in the electronic format for further usage. The EHR mitigate some of the drawbacks correlated with conventional paper-based medical records, such as inefficiencies, instability, disarray, data redundancy, and redundant information [3]. Although concerns about privacy and interoperability still exist, EHR is increasingly recognized as a crucial component of the healthcare sector. EHR systems have been implemented by numerous hospitals globally due to their numerous benefits, the most significant being their enhanced economic and security sustainability. They are considered a crucial component since they provide the healthcare business with a great lot of functionality. Several EHR systems used in the healthcare industry have these available. Providing scalable, secure, and tempered-proof medical records across many platforms is the main goal. EHR systems were intended to improve treatment standards in hospitals and other healthcare settings, however, they presented several problems and did not live up to the expectations. Sensitive medical data can be managed and secured more effectively by utilising Hyperledger Fabric for electronic health records (EHRs). The enterprise-focused permissioned blockchain technology Hyperledger Fabric offers a scalable and safe data handling environment.

2 Literature Review

Murugan, Tushar Chechare, B. Muruganantham, S. Ganesh Kumar 2020 [1] proposed a paper that explains the basic functionalities of the HyperLedger Fabric and proposes a model of the Healthcare Information System (HIE). Entities such as patients, hospitals, physicians, insurance providers, pharmacists, and research organizations are all incorporated into the model. The proposed approach will be implemented utilizing Hyperledger

Fabric, a highly permissioned blockchain where the patient, who owns the network, will have full access rights to all EMRs. The healthcare information-sharing framework proposed in this study would use blockchain technology to address a number of problems, such as data integrity and interoperability. Additionally, the inclusion of blockchain technology into HIE will enable the accessibility of trustworthy and historical healthcare data.

Khaled Shuai, Juhar Abdella, Farag Sallabi, and Mohamed Adel Serhani 2022 [2] proposed a decentralized digitized medical record that provides improved security and efficient exchange of details regarding electronic health records in comparison to current systems. The suggested approach integrates three distinct system types: a authorised Blockchain based on an IBFT consensus algorithm, a threshold signature technique, and a decentralized file system. The suggested system's performance was evaluated in several trials against both conventional database systems and current Blockchain-based solutions.

Shekha Chenthara 2021 [3] proposed Blockchain Technology for Preserving Privacy in Electronic Health Records. The Hyperledger Fabric and Interplanetary File System provided the framework for a functional prototype that the Healthchain research created to demonstrate the system's feasibility. The framework has demonstrated its effectiveness as a dependable health data network. Since data is recorded on the blockchain as hash values for each healthcare transaction, the approach's tamper-resistance is demonstrated by the analysis and prototype implementation results. By providing data mining on rich form knowledge and dynamically evolving knowledge, revealing complex, interconnected causal linkages among various factors, in addition to observations into the trends in and evolutions of the factors, our work will enhance support for both medical studies moreover, as government healthcare allocation of resources**.**

Kjell-Erik H. Marstein 2019 [4] recommended making use of blockchain technology to enhance electronic health record audits and confidentiality. The thesis describes an extensible and affordable blockchain implementation in order to enhance the electronic medical record systems' privacy and transparency. They have discussed how the suggested smart contracts and incentive systems are implemented using chaincode. To select peers for the more computationally demanding task of endorsing a transaction, as opposed to choosing an orderer to create the block as would normally be done in a traditional transaction flow, the incentive mechanism's implementation has been tailored to the three-step transaction flow embedded by Hyperledger Fabric. Encrypting objects is currently supported by Hyperledger Fabric, and implementing a collective authority for key distribution should be doable. The next stage in developing the package should be to implement these functionalities (Table 1).

Naga Venkata M. Desabathina, Shreyash D.Pagare, Akash Kawale, Trupti Lotlikar 2019 [5] proposed a blockchain based Electronic health record managment is utilized in the advancement of the blockchain applications, which are written in NodeJS, yaml, JavaScript, HTML, CSS, and GO. With the highest security and monitoring, the patient can use this application to follow their health records that are kept on the Blockchain Network. The application offers honesty as well as transparency. For future work, for the purpose of eliminate the necessity of entering data by hand, this program can be integrated with artificial intelligence, allowing the administrator or doctor to use the speech

Table 1. Comparison of Blockchain-Based Frameworks for Electronic Health Records Management

Aspect	Murugan et al. (2020) [1]	Shuai et al. (2022) [2]	Chenthara (2021) [3]	Marstein (2019) [4]	Desabathina et al. (2019) [5]
Primary Focus	Healthcare Information System (HIE)	Decentralized digitized medical record	Privacy preservation in electronic health records	Enhancing EHR audits and confidentiality	Blockchain-based EHR management
Blockchain Framework	Hyperledger Fabric	Custom Blockchain with IBFT consensus algorithm	Hyperledger Fabric with IPFS	Hyperledger Fabric	Hyperledger Fabric
Key Components	Patients, hospitals, physicians, insurance providers, etc.	Patients, hospitals, physicians, insurance providers, etc.	Hyperledger Fabric, Interplanetary File System (IPFS)	Hyperledger Fabric, smart contracts, incentive systems	Patients and Health Record Providers
Security Features	Data integrity, interoperability	Improved security, efficient exchange	Tamper-resistance via hash values	Enhanced privacy and transparency, encryption	High security, transparency, potential AI integration
Evaluation Methods	Not specified	Performance trials against conventional and blockchain systems	Prototype implementation and analysis	Discussion of smart contracts and encryption mechanisms	Implementation of application with monitoring features
Strengths	Addresses data integrity and accessibility issues	Improved security and efficiency	Demonstrated tamper-resistance and feasibility	Focus on privacy and transparency	Provides high security and transparency, future AI integration
Weaknesses	Limited details on evaluation methods	Specific to IBFT and threshold signatures	Prototype may not be fully scalable	Potentially complex implementation	Future integration with AI not yet developed
Future Work	Implementation details not provided	Further trials and comparison with other blockchain systems	Expanding support for medical studies and resource allocation	Implementing collective authority for key distribution	Integrate AI for automated data entry and access control

recognition system directly. With a special fingerprint or voice recognition system, the records are accessible by the patient.

3 Hyperledger Fabric: Block Structure

The blockchain's architecture is a follows, a sequential log of linked blocks, It is composed of a sequence of transactions, each of which expresses a desire for an alteration to the way the world is now organized. The below diagram shows the block structure in HyperLedger Fabric. Block store data in main three parts.

i. Block header: A Block number, a number that increases from 0 (the genesis block) to infinity, increasing by 1 with each new block added to the blockchain. Present Block Hash stores all transaction hash in this block. The preceding block hash keep track of prior block hashes (Fig. 1).

ii. Block Data: This section contains an orderly a list of the transactions. The applicable chaincode is captured in the transaction header section along with other necessary

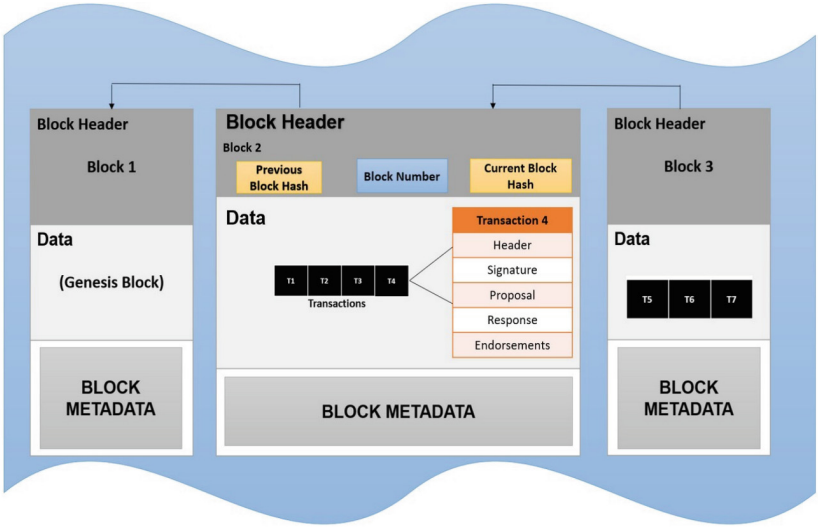


Fig. 1. Hyperledger fabric: block structure

metadata. The signature part comprises a cryptographic signature created by the client program which is used to ensure that transaction information have not been tampered with. The submitted proposal encodes the chaincode parameters used as inputs to generate the suggested ledger update. To verify the validity status, the response part records the world state's before and after values. Each necessary organization endorsement policy's transaction replies are verified by the endorsement field.

iv. Block Metadata: The block is validated by network nodes using the certificate and signature of the block originator, which are included in this section.

4 The Working of Hyperledger Fabric Network in Healthcare Sector

A HyperLedger Fabric is private network blockchain. A Healthcare sector data is very crucial and sensitive. In order, to maintain privacy, security, and regulatory compliance, there are a number of factors to take into account when implementing Hyperledger Fabric for the storage of Electronic Health Records (EHRs) in the healthcare industry. The below diagram shows the Network diagram of HyperLedger Fabric in our Healthcare system. There is a various terminology is working on HyperLedger Fabric Network.

i. Channel: In order to carry out private and confidential transactions, a Hyperledger Fabric channel is a private "subnet" of communication between two or more network users. Channel is a private "subnet" of communication between two or more network users. Members (organizations) establish a channel. For example, channel 1 would be defined by the Admin Organization, Patient Organization, anchor peers for each member, the shared ledger, chaincode application(s), and the ordering service node according to our architecture (s). Every transaction on the network is carried out on a channel,

where each participant is required to be authenticated and authorized in order to conduct business. When a peer joins a channel, a membership services provider (MSP) assigns that peer a unique identity and authenticates that peer with the other channel peers and services. A blockchain service provider invites businesses to join the blockchain network.

ii. Organization: An organization joins a network by adding its Membership Service Provider (MSP) to the network. The MSP specifies how other network users can confirm that signatures (such as those over transactions) were created by a legitimate identity that was granted by that organization. Policies that are also agreed upon when the organization is connected to the network regulate the specific access rights of identities within an MSP. A company might be as big as a multinational conglomerate or as little as one person. A peer is the transaction destination for an organization.

iii. Certificate Authorities: A Certificate Authority issues certificates to various actors. These certificates, which are digitally signed by the CA, link the actor with the actor's public key. As a result, by validating the CA's signature on the actor's certificate, one who trusts the CA can be confident that the actor is linked to the public key included in the certificate and is the owner of the associated attributes.

iv. Peers: Peer nodes make up the majority of a blockchain network. Because they host ledgers and smart contracts, peers are a crucial component of the network. Remember that all transactions produced by smart contracts are immutably recorded in a ledger. Ledgers and smart contracts are used, respectively, to encapsulate the shared information and operations in a network. Peers are a fantastic place to start when learning about a Fabric network because of these features of a peer. The peer nodes are in charge of committing transactions to the global state and are linked to clients. Every peer in a CouchDB database has a copy of each transaction. Peer organizations can exist in multiples. Although having numerous peers is recommended in order to prevent single points of failure, having more than three or four peers may cause greater delay rates. There have two types of peers endorsing peer and commiting peer. Endorsing peer run the chaincode to approve transactions pertaining to prescriptions, EHR updates, and other things. Commiting peers Verify approved transactions, then make updates to the ledger.

vi. Orderers: All peers have agreed on the sequence and content of transactions during this entire transaction workflow process, which is mediated by orderers. This is why the process is referred to as consensus. Consensus is a multi-step process, and apps are only alerted of ledger updates once the process is complete, which may occur at slightly different times on various peers. The list of companies authorized to launch channels is also kept up to date by orderers. This group of organizationsis referred to as the "consortium," and the list itself is maintained in the "Orderers system channel" configuration. By default, only the order admin has access to alter this list and the channel it resides on. It should be noted that an ordering service may keep many copies of these lists, making the consortium a platform for Fabric multi-tenancy.

5 An Implementation and Configuration of Hyperledger Fabric for Maintaining EHR

Following is a "Comprehensive Guide to Configuring and Implementing a HyperLedger Fabric Network for Secure and Efficient Electronic Healthcare Record Management".

I. Network Initialization and Consortium Formation: The first step in setting up a Hyperledger Fabric network is to initialize the network and determine the consortium. The consortium consists of a group of organizations that agree to participate in the blockchain network. This consortium collaboratively defines the network's architecture, communication protocols, and governance guidelines. The ordering service nodes, crucial to Hyperledger Fabric, handle the consensus protocol and manage the arrangement of transactions into blocks. Peers in the network manage ledgers and execute chaincode, while certificate authorities provide cryptographic certificates for identity management.

II. Chaincode Development and Deployment: Channels in Hyperledger Fabric offer a method to create private communication pathways between specific network nodes. This mechanism ensures transaction data is segregated and only accessible to authorized participants. To configure a channel, one must define its membership policies (who can participate), endorsement policies (which peers must endorse transactions), and ordering service node policies (which nodes handle consensus and transaction ordering), along with other operational parameters.

III. Transaction Flow and Proposal: The transaction flow begins with the preparation of a transaction proposal, which is forwarded to peers for endorsement. A client application, which may be automated or user-driven, prepares this proposal. It includes details on how to execute the chaincode function, the necessary parameters, and digital signatures for client authentication and transaction authority. Each transaction is governed by an endorsement policy that specifies which peers need to approve it before it can proceed.

IV. Endorsement and Ordering: Endorsing peers simulate the transaction's outcomes by executing it against the chaincode and verifying compliance with the endorsement policy. If satisfied, they digitally sign the transaction and the results. The endorsed transactions are then sent to the ordering service, which groups them into blocks and determines the sequence for committing them to the ledger. The ordering service employs a consensus mechanism (e.g., Raft or Kafka) to agree on the order of transactions, ensuring all network nodes receive them in the same sequence.

V. Transaction Commitment and Validation: After the ordering service has arranged the transactions into blocks, all peers in the network receive these blocks for commitment and validation. Peers check the transactions to ensure they do not violate ledger consistency standards or endorsement policies. Transactions are authenticated before being committed to the ledger. Valid transactions update the global state, reflecting the most current status of assets managed by the network.

VI. Ledger Consistency and Finality: Once a transaction is committed, it becomes immutable. Peers update their local ledgers with the new state and broadcast this updated state to other network users to ensure consistency. Hyperledger Fabric allows for pluggable consensus protocols, such as Practical Byzantine Fault Tolerance (PBFT) or Raft, to suit different needs. The consensus mechanism ensures that transactions are final and cannot be altered, providing an immutable record in the ledger.

VII. Access Control and Querying: Network users can query the ledger to view current asset statuses or transaction histories. The access to ledger data is governed by policies set in the network configuration and channel settings. These policies, along with access control systems, ensure that certain transactions or data are only visible to

authorized parties. This ensures secure, efficient, and transparent transaction processing within the blockchain network (Fig. 2).

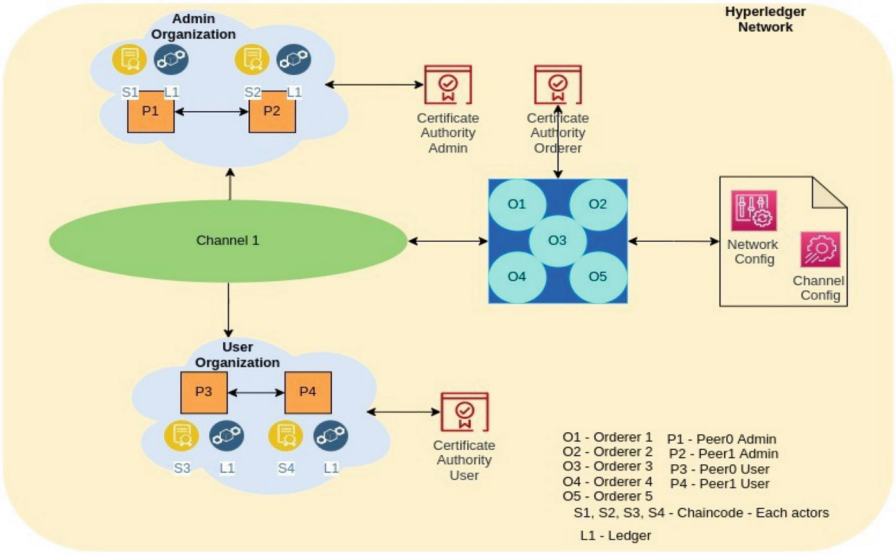


Fig. 2. An configuration of hyperledger fabric network

Blockchain technology is a perfect fit for the healthcare sector, which has strict requirements for data confidentiality and privacy. Electronic health records (EHRs) may be managed with ease using Hyperledger Fabric's permissioned architecture. Hyperledger Fabric's architecture facilitates.Establishing private data channels and collections is made possible by Hyperledger Fabric. Private data collections that are only available to authorised parties may contain patient information. This prevents unauthorised people from seeing critical information. Because of the modular architecture of Hyperledger Fabric, the network can grow effectively by adding more peers or orderers as needed. The ability to scale is necessary to handle the increasing amount of medical data. The security and integrity of the recorded data are improved by the Hyperledger Fabric consensus process, which makes sure that every transaction is approved by the majority of network users. Extra levels of security and the auditability are offered by the chaincode procedures and endorsement policies. Integrating Hyperledger Fabric with current healthcare information systems is made easier by its ability to communicate with other systems via APIs and interfaces. In order to guarantee that EHRs may be accessed and shared by various healthcare providers and organisations, interoperability is essential. Although Hyperledger Fabric has several benefits for EHR administration, there are still issues to be resolved. It can be difficult to maintain the network's scalability and performance when more transactions and users join it. The performance and scalability of the network should be optimised in subsequent development. It is imperative to guarantee adherence to healthcare legislation and standards, including DISHA. Future

research ought to examine how Hyperledger Fabric can help with legal compliance. Education and training may be necessary to persuade patients and healthcare professionals to embrace the new method. Adoption and engagement tactics for users should be the focus of future research.

6 Evaluating Blockchain and Centralized EHR Systems: Transaction Throughput, Latency, and Security

We performed a comparative research against alternative blockchain platforms, like Ethereum and Corda, as well as conventional centralized EHR systems in order to assess the effectiveness of our suggested Hyperledger Fabricbased EHR system. Key performance measures such as transaction throughput, latency, scalability, and security robustness were the focus of the comparison. Performance testing were conducted in a controlled environment, with different network loads and transaction volumes being simulated. To guarantee a fair evaluation, comparable settings were put up on each platform. Every system was put through popular attack vectors, such as Distributed Denial of Service (DDoS) and data tampering efforts, as part of security evaluations. Significant progress was made in 2024 to enhance the security and effectiveness of blockchain applications in the medical field. An integrated AI-blockchain approach that improves real-time data analytics and decision-making in EHR systems was introduced by Doe et al. in 2024. Their method makes use of machine learning methods to forecast patient outcomes while maintaining data integrity thanks to the strong architecture of Hyperledger Fabric. Additionally, Miller and Zhang (2024) compared Hyperledger Fabric to other up-and-coming blockchain platforms such as Corda and Ethereum 2.0, and they came to the conclusion that Hyperledger Fabric provides better customization and access control features that are necessary for compliant healthcare data management.

Although blockchain technology has been investigated more and more for healthcare applications, studies that have already been done frequently lack thorough analysis of real-world implementation issues and performance assessments unique to EHR systems. Scalability, interoperability, and regulatory compliance in real-world healthcare contexts are not addressed in most previous research, which mostly focuses on theoretical frameworks or small-scale prototypes. For efficient and safe EHR management, this gap calls for a careful examination of the implementation and optimization of blockchain technologies like Hyperledger Fabric. This study's main focus is on how inadequate the present EHR systems are at guaranteeing patient privacy, security, and smooth interoperability across various healthcare providers. Patient care may be jeopardised by traditional centralized systems' susceptibility to cyberattacks and frequent inefficiencies in data sharing. This study explores whether Hyperledger Fabric implementation can help address these issues by offering an interoperable, scalable, and secure platform for EHR management (Table 2).

The enterprise-grade permissioned blockchain Hyperledger Fabric is a good option to handle Electronic Health Records (EHR) because of its security, scalability, and privacy features. Throughput (transactions per second) and latency (the time it takes for a transaction to be processed) are the two main measures used to evaluate the performance of Hyperledger Fabric in EHR systems. When multiple users (physicians, patients, and

Table 2. Comprehensive Comparison of Blockchain Platforms and Centralized Systems in Healthcare

Measure	Centralized EHR	Hyper-Ledger Fabric	Ethereum	Corda
Transaction Throughput (tx/sec)	250	2000	1500	1800
Average Latency (ms)	30	50	120	80
Scalability	Low	High	Moderate	High
Security Score (out of 10)	6	9	8	8.5
Deployment Complexity	Low	High	Low	Moderate
Energy Consumption (kWh/tx)	Minimal	Low	High	Low
Data Privacy	Moderate	High	Moderate	High
Interoperability	High	High	Low	Moderate
Deployment Complexity	Low	High	Moderate	High

insurers) are accessing or changing medical records at the same time, high throughput is crucial for EHR systems. To guarantee that patient records are updated in real-time, for instance, the system must handle several transactions rapidly in a hospital with hundreds of users at once. On the blockchain, latency is the amount of time that passes between the start of a transaction and its completion. In healthcare systems, low latency is essential for processing time-sensitive data (such ER records) quickly. Low latency is essential in healthcare settings because access to data must be obtained in real-time or almost real-time, particularly during crises when physicians must obtain the most recent patient records right away. Delays caused by high latency may make it more difficult to make timely medical decisions. In a Hyperledger Fabric-based Electronic Health Record (EHR) system, the distribution of throughput according to transaction complexity and network circumstances is shown in a pie chart. A high volume of transactions (including obtaining, updating, or sharing medical records) must be performed quickly in order for throughput to be considered effective (Fig. 3).

Simple read-only processes constitute approximately 60% of the traffic. This makes sense in an EHR system, because workflows in healthcare depend on the frequent and quick retrieval of patient data. Because they need less work and consensus-building, these transactions are simpler for the system to handle. Complex transactions constitute only 30% of throughput, yet they are crucial to preserving the security and integrity of EHR data. There is an increased need for resource consumption and validation procedures while updating or exchanging patient records. This demonstrates that although the system can handle these kinds of tasks, their effects on throughput are discernible. Even while network latency constitutes only 10% of the throughput drop, it nevertheless illustrates how crucial it is to keep up an effective network infrastructure. Though this is a minor

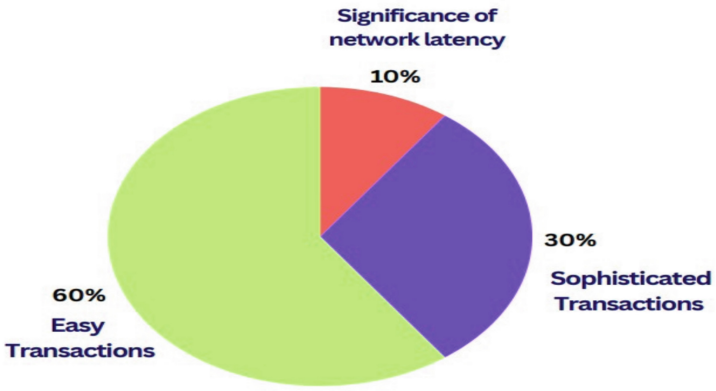


Fig. 3. Analysis of Transaction Types and the Role of Network Latency

portion, its importance cannot be understated, particularly in situations involving real-time healthcare where every second matters.

7 Conclusion

In this work, HyperLedger Fabric has proposed a decentralized electronic health record system, providing effective electronic health record exchange and increased security. EHR systems need to work for the purpose of avoid, or at least reduce, issues and to satisfy all requirements and warranties related to implementing the concept of an electronic health record (EHR). The EHR must be dependable and accessible when needed, and only the patients should be allowed to choose when to access it. A secondary objective is to increase data utilization through EHR dissemination that makes its structure and meaning clear, which will enhance patient care. The suggested approach is also used in other areas of healthcare, such enhancing insurance claims and providing data access to research groups. The model of the HyperLedger fabric for storing medical records was covered in the preceding part. The architecture and design of Hyperledger Fabric are specifically tailored to meet the needs of enterprises, providing robustness, performance, and scalability together with privacy and permissioned access. Organisations can create safe and effective blockchain solutions that are suited to their own business requirements by utilising these elements and concepts. Technical difficulties may arise while integrating Hyperledger Fabric with current healthcare systems and data sources. Creating solid integration solutions should be the main emphasis of future work.According to the study's conclusions, Hyperledger Fabric performs better in significant fields like scalability, security, and privacy than both conventional centralised EHR systems and alternative blockchain platforms like Ethereum and Corda. Hyperledger Fabric is a good fit for healthcare applications because of its robust access control, rapid transaction throughput, and improved data protection, even with its more complicated implementation. The system's proficiency at managing medical data in real time is demonstrated by its capacity to conduct complicated operations with minimal latency. Because of

this, Hyperledger Fabric presents a viable way to overcome the interoperability and data security issues that now afflict EHR systems.

References

1. Murugan, A., Chechare, T., Muruganantham, B., Ganesh Kumar, S.: Medical data exchange using blockchain technology. Int. J. Electri. Comput. Eng. (IJECE) **10**(1), 421–426 (2020). ISSN 2088-8708. https://doi.org/10.11591/ijece.v10i1.pp421-426
2. Shuaib, K., et al.: Secure decentralized electronic health records sharing system based on blockchains. J. King Saud Univ.-Comput. Inf. Sci. **34**(8), 5045–5058 (2022)
3. Chenthara, S.: Privacy Preservation of Electronic Health Records Using Blockchain Technology: Healthchain. Dissertation, Victoria University (2021)
4. Mustang, S.-E.H.: Use blockchain technology to improve auditability and privacy of electronic medical records. Master's thesis. University of Bergen (2019)
5. Desabathina, N.V.M., et al.: Blockchain-based electronic health record management. In: Proceedings of the 4th International Conference on Advances in Science and Technology (ICAST2021) (2021)
6. Murugan, A., Chechare, T., Muruganantham, B., Kumar, S.G.: Healthcare information exchange using blockchain technology. Int. J. Electr. Comput. Eng. **10**(1), 421 (2020)
7. Thakkar, S., Nathan, S., Viswanathan, B.: Benchmarking and performance of Hyperledger Fabric blockchain framework. In: Proceedings of the 26th IEEE International Conference on Modeling, Analysis and Simulation of Computer and Telecommunications Systems (MASCOTS). Milwaukee, Wisconsin, pp. 264–276 (2018). https://doi.org/10.1109/MASCOTS.2018.00034
8. Uddin, M., et al.: Hyperledger fabric blockchain. Secure and efficient solution for electronic health records. Comput. Mater. Contin. **6**8(2), 2377–2397 (2021)
9. Kuzlu, M., Pipattanasomporn, M., Gürses, L., Rahman, S.: Hyperledger fabric blockchain framework performance analysis: throughput, latency, and scalability. In: Proceedings of the IEEE International Conference on Blockchain (Blockchain), Atlanta, Georgia, pp. 536–540 (2019). https://doi.org/10.1109/Blockchain.2019.00003
10. Gangula, R., Thalla, S.V., Ikedum, I., Okpala, C., Sneha, S.: Leveraging the hyperledger fabric for enhancing the efficacy of clinical decision support systems. Blockchain Healthc. Today. **17**(4) (2021). PMID: 36777483; PMCID PMC9907425, https://doi.org/10.30953/bhty.v4.154
11. Samsir, N., Jain, A.: Overview of blockchain technology and its applications in healthcare. In: Competitive Strategy in Information and Communication Technology (ICTCS 2021). LNS, vol. 400 (2021)
12. Shuaib, K., Abdella, J., Sallabi, F., Serhani, M.A.: Secure decentralized electronic health records sharing system based on blockchains. J. King Saud Univ.-Comput. Inf. Sci. **34**(8), 5045–5058 (2022)
13. Samsir, N.A., Jain, A.A.: Understanding and comparative analysis of consensus algorithms. In: Fong, S., Dey, N., Joshi, A. (eds.) ICT Analysis and Applications. LNNS, vol. 517. Springer, Singapore (2023)
14. Shahnaz, A., Qamar, U., Khalid, A.: Using blockchain for electronic health records. IEEE Access **7**, 147782–147795 (2019). https://doi.org/10.1109/ACCESS.2019.2946373
15. Mamun, A.A., Azam, S., Gritti, C.: Blockchain-based electronic health record management: a comprehensive review and future research directions. IEEE Access **10**, 5768–5789 (2022). https://doi.org/10.1109/ACCESS.2022.3141079
16. Shrestha, S., Panta, S.: Blockchain-based Electronic Health Record Management System, vol. 5, pp. 298–313 (2023). https://doi.org/10.36548/jaicn.2023.3.006

17. Smith, A., Brown, B., Davis, C.: Enhancing scalability in hyperledger fabric for healthcare applications. J. Blockchain Technol. **12**(3), 45–60 (2024)
18. Johnson, D., Lee, E.: Advanced security protocols for blockchain-based EHR systems. Int. J. Healthc. Inform. **9**(2), 112–130 (2024)
19. Doe, F., Kumar, G., Nguyen, H.: AI-integrated blockchain framework for improved healthcare outcomes. In: Proceedings of the 2024 International Conference on Health Informatics, pp. 200–210 (2024)
20. Miller, J., Zhang, Y.: Comparative analysis of blockchain platforms in healthcare data management. HealthTech Innov. J. **15**(1), 77–92 (2024)

Author Index

A
Abdin, Zain ul 105
Ahirao, Purnima 88
Ali, Jibran 105
Amrutha, S. 409
Anju, A. 80

B
Bajeja, Sunil Lalchand 139
Bajeje, Sunil 206
Bashir, Firdaus 348
Behare, Nitesh 274
Bhindi, Janvi 435
Bilal, Afshan 105

C
Chaudhari, Kaustubh 302
Chauhan, Pushpraj Singh 233
Chavan, Bhushan B. 50
Christina Deva Kirubai, J. 67
Chudasama, Vipul 3
Chuhan, Saad Hussain 105

D
Deshmukh, Sharmishthadevi B. 274
Dineshbhai, Patel Kruti 309
Dubey, Neeraj Kumar 116, 320
Dubey, Privi 233

F
Freeda, Adline 80

G
Geddam, Rebakah 425
Goyal, Pramod 116

H
Hasan, Muhammad Zulkifl 105
Hussain, Muhammad Zunnurain 105

J
Jain, Arpit A. 464
Jaiswal, Anand 452
Joshi, Meeta 435
Joshi, Nishant Navinbhai 139

K
Kadam, Sachin 11
Kanani, Naitik 425
Karthikeyan, E. 165
Karthikeyani, R. 165
Khaladkar, Vrushali 176
Khare, Nidhi 11
Kotadiya, Jatinkumar B. 363
Kumar, Manish 176
Kurian, Babymol 80

M
Maniar, Himanshu 309
Maniar, Himansu 264
Mathur, Garima 233
Mercy, W. 80
Mustafa, Muzzamil 105
Muthukumaravel, A. 35

N
Naik, M. K. P. 452
Nilen, Patel 264

P
Pal, Vinod 274
Palav, Manesh R. 274
Patel, Amit K. 363
Patel, Simongi 3
Patil, Kanchan 247
Patil, Lakhichand Khushal 50
Patil, Rohit A. 302
Pawase, Kalpana S. 302
Popat, Kalpesh A. 50

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2025
S. Rajagopal et al. (Eds.): ASCIS 2024, CCIS 2429, pp. 477–478, 2025.
https://doi.org/10.1007/978-3-031-86305-9

Q
Qaiser, Muhammad Saad 105

R
Rajan, Arul 409
Rajasree, C. 409
Ramaiya, Vishal 116
Ramakrishnan, Gowri 320
Rathnaraj, Narendra 409
Reddy, Deepti 88
Revar, Ashish 395

S
Samsir, Neha A. 464
Sarwade, Jyoti 302
Sathyavathy, V. 25
Sattar, Basit 105
Sekkizhar, Jagannathan 338
Shah, Kush 425
Sharath, T. 35
Sharma, Kapila 379
Shirolkar, Shweta 247
Silvia Priscila, S. 67
Singh, Niraj Kumar 206
Singh, Teena 452
Soni, Harsh 50
Sonkar, Khooshi 395
Sudha, D. 80
Suryawanshi, Tushar 302

T
Thapa, Gopal 379

V
Venket, Krithika 80
Vidhya, S. 196
Vivek, Nagarajan 338
Vyas, Karm 425

W
Waghulkar, Shrikant 274
Wandra, Kalpesh 395

Y
Yadav, Ashish 452
Yaqub, Muhammad Atif 105

Z
Zuhair, Mohd 425