Studies in Computational Intelligence 1178

Mohammed A. Al-Sharafi
Mostafa Al-Emran
Moamin A. Mahmoud
Ibrahim Arpaci *Editors*

Current and Future Trends on AI Applications

Volume 1

Springer

Studies in Computational Intelligence

Volume 1178

The series "Studies in Computational Intelligence" (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

Indexed by SCOPUS, DBLP, WTI AG (Switzerland), zbMATH, SCImago.

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

Mohammed A. Al-Sharafi · Mostafa Al-Emran ·
Moamin A. Mahmoud · Ibrahim Arpaci
Editors

Current and Future Trends on AI Applications

Volume 1

Springer

Editors
Mohammed A. Al-Sharafi
King Fahd University of Petroleum & Minerals
Dhahran, Saudi Arabia

Moamin A. Mahmoud
Universiti Tenaga Nasional (UNITEN)
Kajang, Selangor, Malaysia

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

Ibrahim Arpaci
Bandırma Onyedi Eylül University
Balıkesir, Türkiye

ISSN 1860-949X ISSN 1860-9503 (electronic)
Studies in Computational Intelligence
ISBN 978-3-031-75090-8 ISBN 978-3-031-75091-5 (eBook)
https://doi.org/10.1007/978-3-031-75091-5

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

Preface

The rapid advancements in Artificial Intelligence (AI) have fundamentally transformed various sectors, redefining the possibilities of technology and innovation. From healthcare and finance to education and entertainment, AI's impact is pervasive and far-reaching. This edited book, *Current and Future Trends on AI Applications*, is a comprehensive exploration of the latest developments in AI and a forward-looking analysis of the emerging trends that will shape the future of this dynamic field.

This edited volume brings together a diverse range of perspectives from experts in the field, covering a wide spectrum of AI applications from disaster management and cybersecurity to healthcare, energy informatics, and beyond. The chapters in this book illustrate the profound impact AI has already made and the even greater potential it holds for the future. The book is structured to provide readers with a thorough understanding of how AI is being applied in various domains today and how it is expected to evolve. Each chapter delves into specific use cases, providing detailed analyses of how AI technologies are being implemented to solve complex problems, improve efficiency, and create new opportunities. This book is intended for researchers, practitioners, and students who are interested in understanding the current state of AI applications and anticipating the future developments that will define the next era of AI-driven innovation. We hope that the insights presented in this book will inspire new ideas, foster collaboration, and contribute to the ongoing dialogue on how to harness the power of AI for the benefit of all.

We would like to express our sincere gratitude to all the authors who contributed their valuable knowledge and expertise to this volume. Their work not only reflects the state-of-the-art in AI but also provides a roadmap for the future, guiding us toward a more intelligent and connected world.

Dhahran, Saudi Arabia — Mohammed A. Al-Sharafi
Dubai, United Arab Emirates — Mostafa Al-Emran
Kajang, Malaysia — Moamin A. Mahmoud
Balıkesir, Türkiye — Ibrahim Arpaci

Contents

The Enhancement of K-Nearest Neighbourhood in Optimizing the Selection of Distribution Centre for Disaster Relief Operation 1
Siti Nabilah Basaran, Waidah Ismail, Mohd Haziqi Hilmi, Rimuljo Hendradi, and Ismail Ahmed Al-Qasem Al-Hadi

Evaluating Advanced Dimensionality Reduction Techniques for Effective Clustering of High-Dimensional Adolescent Development Data 17
Khoula Said Al. Abri and Manjit Singh Sidhu

A Decade-Long Bibliometric Analysis of Electronic Health Record Research Trends (2014–2024) 37
Muhammed Ibrahim, Moamin A. Mahmoud, Mohammed A. Al-Sharafi, and Abba Hassan

Performance Analysis of Deep Learning Models in Detection and Counting of Bees for Hive Monitoring 55
James Alfred Baluran, Dan Lloyd Lomongo, Ivant Fred Ondoy, Jinky Marcelo, Donah Rae Verula, Viceliogems Verula, Mohana Shanmugam, and Lemuel Clark Velasco

Smarter Learning Environment: Exploring the Need for an Intelligent Tutoring System in Iraqi Universities 75
Ahmed Rashid Alkhuwaylidee, Ali khalaf Gatea, and Mohammed F. Alomari

Automated Mandible Segmentation from Computed Tomography Scans Using U-Net and U-Net Transformer 89
Robert William Wacan, Louisse Amadeo Romero, Paul Justine Ardiente, Lemuel Clark Velasco, Mohana Shanmugam, and Chris Jordan Aliac

AI-Enhanced Cybersecurity: A Comprehensive Review of Techniques and Challenges 107
Liyana Adilla binti Burhanuddin, Abdul Samad Bin Shibghatullah, Irma Syarlina Che Ilias, Zanariah Binti Zainudin, and Nurfazrina Binti Mohd Zamry

A Bibliometric Review of Trends and Insights of Internet of Things on Cybersecurity Issues 127
Mushtaq Yousif Alhasnawi, Ahmed Abbas Jasim Al-Hchaimi, Yousif Raad Muhsen, and Amine Lekmiti

Artificial Intelligence Implementation in the Tourism Industry: A Literature Review 149
Weiming Wang, Noorminshah A. Iahad, and Mohammed A. Al-Sharafi

Artificial Intelligence (AI)-Enabled Mobile Fitness Apps and Goal Attainment: Systematic Literature Review 167
Zhang Yanyan, Noorminshah A.Iahad, and Ahmad Fadhil Yusof

Artificial Intelligence in Energy Informatics: An In-Depth Review 183
Safa Oleiwi, Saraswathy Shamini Gunasekaran, Moamin A. Mahmoud, Jaspaljeet Singh Dhillon, Nazirul Nazrin Shahrol Nidzam, and Salama Mostafa

AI Strategies for Handling Disciplinary and Cyber Bullying in Schools 207
Manjit Singh Sidhu and Kirandeep Kaur Sidhu

Artificial Intelligence and Industry 4.0: Catalysts for Sustainable Organizational Development 231
Ala'a M. Al-Momani

Text Classification: A Comprehensive Survey from Traditional Approaches to Deep Learning Methods 247
Ihab L. Hussein Alsammak, Wasan H. Itwee, Moamin A. Mahmoud, and Noor Islam Jasim

Development of Smart Contract Blockchain Integrated Model (Lightblock) for IOT Applications 269
Mathuri Gurunathan, Moamin A. Mahmoud, Ihab L. Hussein Alsammak, and Saraswathy Shamini Gunasekaran

Hybrid Deep Learning for Detecting Hate Speech Across Social Media Platforms 289
Loke Kevin, Chai Ting Yee, Chia Zhi Liang, Stephen Yong, Yeap Cheng Han, Muaadh Mukred, and Fathey Mohammed

AI-Driven Smart City Security and Surveillance System: A Bibliometric Analysis 305
Wong Qi Fan, Aalima Shaura Ismail, Fathey Mohammed, and Muaadh Mukred

Adoption and Implementation Trends of Vehicle-to-Everything (V2X) Technologies: A Comprehensive Bibliometric Analysis 329
Noor Islam Jasim, Saraswathy Shamini, Mohammed A. Al-Sharafi, Moamin A. Mahmoud, Muhammed Ibrahim, and Abba Hassan

Artificial Intelligence Methods and Image Recognition Techniques in Ophthalmic Robotic Surgery: A Review 345
Mukesh Madanan, Saraswathy Shamini Gunasekaran, Moamin A. Mahmoud, Jaspaljeet Singh Dhillon, Salama Mostafa, and Nazirul Nazrin Shahrol Nidzam

Examining Factors Influencing Behavioral Intentions to Use Artificial Intelligence in the Recruitment Processes of United Nations Agencies 365
Noor Al-Sabri, Redhwan Al-amri, and Gamal Alkawsi

A Study on the Use of Enterprise Resource Planning to Improve the Sustainability of Malaysian SMEs 395
Shree Poovan Sivaganthan, Muaadh Mukred, Fathey Mohammed, and Mikkay Wong Ei Leen

Determinants of Biometric as a Service (BaaS) Adoption in the Telecommunication Companies: The Mediating Effect of Perceived Trust 411
Ahmad Abed Al-Hayy Al-Dalaien and Hairoladenan Kasim

The State of the Art in AI-Based Phishing Detection: A Systematic Literature Review 431
Mohammed Alghenaim, Gamal Alkawsi, and Christopher R. Barnhart

The Enhancement of K-Nearest Neighbourhood in Optimizing the Selection of Distribution Centre for Disaster Relief Operation

Siti Nabilah Basaran, Waidah Ismail, Mohd Haziqi Hilmi, Rimuljo Hendradi, and Ismail Ahmed Al-Qasem Al-Hadi

Abstract The abstract should summarize the contents of the paper in short terms, i.e. 150–250 words. Disaster relief operations involve assisting victims in their recovery from either natural or human-caused disasters. In Malaysia, frequent floods pose a recurring challenge, highlighting the importance of efficiently selecting distribution centres for relief efforts, especially given their annual occurrence. A critical aspect of these operations is the strategic choice of operational centres for distributing necessities to victims. However, it is recognized that not all disaster areas can be covered during recovery efforts. To address these challenges, an optimization algorithm is necessary to enhance the coverage of disaster areas. In this context, the K-Nearest Neighbour (KNN) algorithm is employed as a fitness function in both the Genetic Algorithm (GA) and Simulated Annealing (SA). This application aims to optimize the distribution of food from selected distribution centres (DCs). The experiment utilized demand points (DPs) and DCs identified by researchers, adopting a hybrid approach of GA and SA. A performance comparison indicated that the hybrid GA-KNN approach yielded the most optimal solution, with an average fitness value 21% lower than that of the hybrid SA-KNN. This study significantly contributes to the identification of optimal distribution centre locations, ensuring a nearly equal distribution of DPs in each selected location. Such optimization facilitates efficient real-world aid distribution in disaster areas, addressing both time and cost concerns.

S. N. Basaran · W. Ismail (✉)
Faculty Science and Technology, Universiti Sains Islam Malaysia, 71800 Nilai Negeri Sembilan, Malaysia
e-mail: waidah@usim.edu.my; ismailahmed@ucsiuniversity.edu.my

M. H. Hilmi
Institut Islam Hadhari, Universiti Kebangsaan Malaysia, UKM, 43600 Bangi, Selangor, Malaysia

R. Hendradi
Information System Study Program, Universitas Airlangga, Indonesia Kampus C, Mulyorejo, Surabaya, Indonesia

I. A. A.-Q. Al-Hadi (✉)
Institute of Computer Science and Digital Innovation, UCSI University, Kuala Lumpur, Malaysia
e-mail: waidah@usim.edu.my

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_1

Keywords Distribution centre · KNN · Genetic algorithm · Simulated annealing · Demand point · Distributional effect

1 Introduction

Please one of the challenges faced by humanitarian organizations is the limited use of decision-making technologies in aiding disaster relief operations. Malaysia experiences flood issues almost every year [1]. During flood disasters, the primary objective is to ensure an organized and effective provision of assistance and aid to the affected individuals, starting from the national level downwards [2]. This approach primarily responds to flood disasters [3]. Implementing strategic selection principles for distribution centres can facilitate volunteers in gathering and distributing essential items [4, 5]. Malaysia has experienced several major flood catastrophes since 2015 [6]. Issues such as deforestation and unmanaged land use have been prevalent causes of flooding across the country [7]. Every year, thousands of people are relocated to evacuation centres when areas are ravaged by floods, though few lives are lost as a result of these calamities. The floods have caused massive losses and devastation, with thousands of homes destroyed and millions of ringgits in property lost [8].

Floods are natural geohazards that occur when heavy rains persist over an extended period. In Malaysia, floods are classified into two types: monsoon floods and flash floods. Monsoon floods are caused by the Northeast Monsoon season, which begins in early November and ends in March, bringing significant rainfall to the east coast states of Peninsular Malaysia and western Sarawak. Flash floods typically occur in rapidly developing areas due to a sudden rise in water levels, high velocity, and large amounts of debris [9]. Small-scale flash flood crises are climate-related disasters that can threaten various components of the system, particularly in densely populated areas such as Kuala Lumpur, Malaysia [10]. Losses and damages from flood disasters are also increasing, while non-economic losses are rarely fully accounted for during flood risk assessment and management [11]. For example, the recent floods in Selangor, Malaysia, from December 18 to 22, 2021, were estimated to have caused at least RM 20 billion in damage [12].

Traditionally, disaster management in Malaysia follows a government-centric top-down approach. Inefficiencies and challenges in distributing basic needs impact the victims, leaving them without necessary support during flood tragedies. Efficient disaster management can ensure timely assistance to victims in all aspects [13].

Climate change can greatly influence the occurrence of natural disasters such as floods, landslides, tornadoes, and earthquakes. A key concern is the rising frequency of floods, driven by altered weather patterns. For example, in the Philippines, heavy rainfall has affected the southern island of Mindanao and later spread to the central region of Visayas. The National Disaster Risk Reduction and Management Council (NDRRMC) reported that on the same day, 5,119 people were displaced to evacuation camps in the Caraga Region, with another 345 in the Davao region [14].

In Malaysia, a flood on October 5, 2020, impacted around 60 villages in the districts of Kota Kinabalu, Kota Belud, and Tuaran. Many homes, health facilities, and public buildings were inundated. The Sabah State Information Department set up six relief centres, providing shelter for 339 people from 136 families who had been evacuated [15]. Furthermore, additional flooding in parts of Peninsular Malaysia forced over 8,000 people to leave their homes [16]. The heavy rains were linked to the annual monsoon season, a common occurrence in Malaysia. According to data from Malaysia's National Disaster Command Centre (NDCC), as of December 20, 7,781 people had evacuated in Terengganu, 1,454 in Kelantan, and 38 in Pahang.

To address this issue, the K-Nearest Neighbour (KNN) algorithm is used to identify the closest distribution centre (DC). The Simulated Annealing (SA) or Genetic Algorithm (GA) is then employed to optimize this nearest location, with KNN functioning as the fitness function for both SA and GA. A comparative analysis of the optimization results shows that the hybrid-GA method produces better outcomes than the hybrid-SA method. This approach is known as memory-based classification because it requires training examples to be stored in memory during runtime [17].

Over the years, the use of optimization techniques in disaster relief operations has expanded significantly [18]. Optimization offers an alternative for addressing complex challenges in disaster relief, including transportation, routing, roadway repair, and issues related to location and allocation [19]. Optimization has proven highly effective; for instance, the hill climbing algorithm is known for its speed and efficiency in extracting low-level features [20]. Both Genetic Algorithms (GA) and Simulated Annealing (SA) are also recognized as effective solution methods [21]. Combining optimization technology with domain expertise can help tackle complex challenges. However, the use of evolutionary algorithms still presents several limitations and drawbacks [18]. Conducting timely and effective relief efforts is essential for maintaining stability in disaster-affected regions [22]. Disasters can lead to long-term consequences by disrupting assets and economic activities [23]. Therefore, adequate external assistance is crucial during the relief stage to support the recovery of impacted communities and to reduce their vulnerability to future disasters.

The inefficiency in selecting distribution centres during disaster relief operations presents challenges for volunteers involved in humanitarian efforts [24]. A key element of disaster relief is the distribution of essential supplies to affected regions, making the strategic selection of operation centres a critical concern. It has been observed that not all disaster-stricken areas can be fully covered during recovery efforts [25] due to the uneven distribution of needs within the affected regions.

There is currently no effective decision model to assist in identifying strategic disaster relief distribution centres that can sufficiently cover all affected victims within a disaster area. This paper introduces a genetic algorithm-based KNN approach to enhance the identification of optimal distribution centre locations and to reduce disparities in demand point coverage across these centres. Additionally, after determining the distribution centre locations on the disaster map, the proposed method will identify the most suitable nearby neighbour nodes.

2 Related Works

2.1 K-Nearest Neighbour

The K-Nearest Neighbour (KNN) algorithm, originally created by Fix and Joseph Hodges in 1951 and later expanded by Thomas Cover for use in classification and regression, is a supervised learning technique. It categorizes new instances based on their proximity to existing neighbours [25]. KNN is highly regarded for its simplicity and ease of use, enhancing the traditional nearest neighbour method [26]. The algorithm works by taking a majority vote among the 'K' nearest neighbours, where 'K' is a small positive integer. Notably, KNN is effective with large datasets and demonstrates resilience to noisy training data.

In their 2020 study, Shahabi et al. introduced new ensemble models that combined bagging as a meta-classifier with KNN to predict flooding in northern Iran, utilizing 10 conditioning factors for flood prediction assessment [27]. During disaster scenarios, humanitarian logistics are crucial for making decisions about location and resource allocation. The researchers used Monte Carlo simulations to generate random numbers and scenarios. KNN has also been effectively used in the classification of blood cell images. Additionally, this study leveraged Google Maps to simulate areas prone to flooding.

2.2 Heuristic Search

When distributing food to flood victims, the primary concerns are time and cost. Optimization is essential to ensure that volunteers can fully cover the affected area. Heuristics are often used to approximate effective solutions in large-scale problems. Although heuristics may not always provide the absolute best solution, they are designed to address real-world challenges efficiently [28]. The fitness function plays a crucial role in heuristic searches. In this paper, we implemented two optimization algorithms, namely SA and GA, while utilizing the KNN algorithm for the fitness function.

Ismail et al. [29] conducted a comparison between SA and Hill Climbing (HC). Subsequently, this study was extended to include SA and GA, specifically in the realm of blood cell image processing [29]. The comparison revealed that SA outperformed GA in accurately detecting blood cells, while GA faced challenges due to overlapping cells [30].

In the domain of real-time Emergency Medical Service (EMS), SA was applied to minimize total travel time for patients transported to the hospital during emergencies. The results demonstrated that SA provided high-quality solutions while requiring less CPU time. As a result, SA is considered suitable for real-time support systems [31].

GA draws inspiration from the biological evolution process [32], reflecting the Darwinian principle of natural selection and survival of the fittest in optimization.

Operating with a fixed-size population, GA employs genetic operations like crossover and mutation to evolve solutions, selecting the most suitable offspring for propagation to subsequent generations [18].

As a global optimization search algorithm, GA possesses numerous unique properties compared to others. It has demonstrated success in various applications, including combinatorial optimization, machine learning, signal processing, adaptive control, artificial life, and many other fields.

The identification of optimal shelter locations for flood evacuation planning was addressed in a study by [33]. They approached the issue by constructing a bi-level programming model, with the upper level dedicated to location determination and the lower level focused on understanding travelers' behaviors. Their methodology involved the implementation of a Genetic Algorithm (GA) that employed genetic operations to evolve a population of complete solutions. A double-stage algorithm was utilized to assess the fitness value of each chromosome.

Another approach involves optimizing the location of catastrophic rescue centres. Tzeng and Chen [34] introduced a model that accounts for the probability of disaster occurrence, the disaster diffusion function, and the rescue function [34]. They developed a Genetic Algorithm (GA) utilizing binary encoding and incorporated a greedy heuristic for fitness evaluation. Experiments on real-world scenarios showed that the algorithm produced satisfactory results.

In 2007, Yang et al. proposed a maximal covering problem model focusing on the location of medical supply facilities [35]. This model aimed to tackle demand uncertainty and insufficiency in medical supplies by ensuring each demand point received services from multiple facilities located at varying quality levels or distances. To address the location problem, three heuristics were developed, including GA. However, experimental results indicated that while the GA was suitable for resolving small problems, it yielded lower-quality solutions for larger problems, primarily due to premature convergence.

To determine the optimal number and locations of fire stations at an international airport, [36] proposed a fuzzy multi-objective model. The objectives include minimizing the total setup cost and minimizing the longest distance from the fire stations to any incident point. These objectives are converted into a single unified 'min–max' goal, and GA is employed to effectively solve the problem. Efforts were made to extend the model so that different risk categories and obstacles within a given region are considered in the objectives and constraints. GA was adapted by embedding the constraints into the fitness function [37].

After 4 years of studying a highway traffic emergency facility location problem, a heuristic genetic algorithm was proposed. This algorithm uses an n-dimensional 0–1 integer vector to represent chromosomes and employs a heuristic operator to enhance unreliable chromosomes by correcting chromosomes that do not cover all demands and eliminating redundancy in chromosomes that cover locations redundantly [38]. The method has been successfully applied to problems in the highway network of Nanjing, China.

Han and Zhang investigated an extended emergency facility location problem modeled as an integer programming (IP) problem. They proposed GA where each chromosome comprises the serial numbers of each target's emergency facility. Population diversity is maintained by utilizing an adjustable mutation probability. Experiments demonstrate that GA solves the problem much more efficiently than other simple heuristics [37].

Genetic Algorithms (GA) can be applied in image processing, particularly in food distribution alignment, where they are known for enhancing image quality and contrast [39]. Hybrid GA, combined with fuzzy logic, is used to eliminate image noise, such as haze, fog, and smog [39]. Additionally, Hiassat et al. (2017) focused their research on warehouse and inventory location models, simplifying the process of locating these facilities [40].

3 Data Sets

To conduct the research, we assume a post-disaster site where a set of houses belonging to disaster victims serves as the population demand points, and potential distribution centres serve as the target points. The coordinates of each demand points and distribution centre (DC) are randomly chosen. This research specifically focuses on using Google Maps in Kuala Krai, Kelantan. The map data will be extracted into Excel using Visual Basic for Applications (VBA) to obtain the coordinates, as shown in Fig. 1.

The demand points (DPs) are marked within a specified area of N = L × L, where L represents the side length of the map. Coordinates will be retrieved from the marked DPs on the map and used for calculations in determining the k-nearest neighbours. The following assumptions are considered regarding the distribution centre model:

Fig. 1 Displays the location of DCs (green color) and DPs (red color)

(i) Consider Demand Point (DP) as the random distribution of DPs across a given area, N. Each demand point represents a disaster victim seeking assistance, whether it's funds, basic needs, or support from volunteers.
(ii) Designate DC as the potential distribution centre, some of which tend to be optimal distribution centres. These distribution centres serve as places where volunteers gather to distribute aid to those in need at each demand point.
(iii) Let M represent the total number of DPs covered by the DC.
(iv) Define Cost as the value aiming to minimize the difference in the coverage of DPs by the DC.

As for the constraints, we assume that each DP can be covered by its nearest DC, and each DC must be capable of serving a minimum of 50 DPs. We also assume that each DC has its organization managing volunteers for the task, and the capacity of each DC is determined by the organization, adhering to standard procedures and regulations.

4 Methodology

The first step is to retrieve the map to obtain the coordinates. The map will be extracted into Excel using Visual Basic for Applications (VBA) to acquire these coordinates from Google Maps. Using the map grid algorithm, the researcher identifies the positions of DPs and DCs on the map. This method ensures high accuracy and prevents the omission of any DPs in the area. The selection was done randomly.

Initially, users are prompted to mark the roof and potential DC locations after inserting the map. Coordinates are then generated in an Excel table and used in KNN classification via the Spyder scientific Python development tool. Finally, a hybrid KNN approach, integrating both SA and GA, is applied. Figure 2 shows the experimental framework of this study.

4.1 K-Nearest Neighbours

The Euclidean Distance (ED) will be used in KNN to measure the straight-line distance between two points in the Euclidean plane, represented by (x1, y1) and (x2, y2), calculated using Eq. 1.

$$ED(x_1, y_1, x_2, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2} \tag{1}$$

When categorizing new model data, the K-NN algorithm computes the distance values between the target record and other records, ordering these records based on their distances. In the final step of this method, the first k-record is selected from

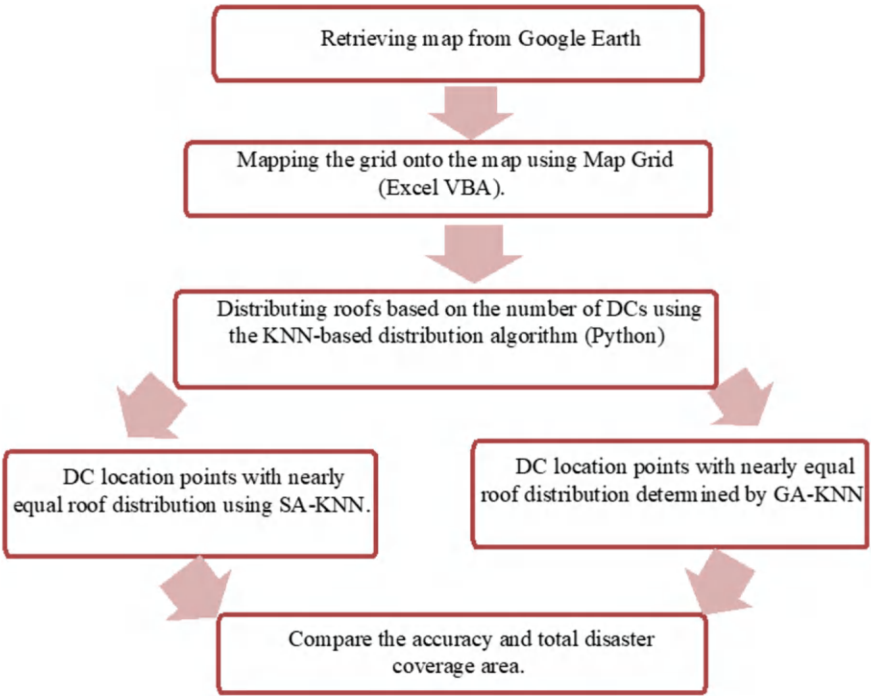


Fig. 2 Framework of experiment

the ordered list as the K-NN. The following pseudocode outlines the typical steps involved in applying K-NN.

Algorithm 1. Pseudo Code of K-Nearest Neighbours Algorithm

Input: *Matrix Features Coordinates*
Output: *A set of neighbours nearest Distribution centre*
Steps:

- *Choose the target Distribution Centre.*
- *Calculate the similarity distance between the target User and other users using Euclidean Distance, as shown in Equation (1).*
- *Sort the users based on their distance values, from the lowest to the highest (representing highest to lowest similarity).*
- *Select k users from the beginning of the sorted list.*

The design and implementation of SA and GA based KNN for estimating the population of disaster victim demand points. In the SA and GA implementation, a fitness function is defined in Eq. 2,

$$f(n_i) = min\left|dc_{n_i} - dc_{n_{+1}}\right| \tag{2}$$

where dc_n represents the set of neighbours of the distribution centre in a population size, dc_{n_i} signifies the distribution centre position indexed by n_i, and dc_{n+1} denotes the subsequent distribution centre position indexed by $n + 1$. Each potential DC is supposed to cover a more equal distribution of DPs and have the closest gap with the mean. After selecting the potentially optimal k parameter, representing the number of nearest neighbours of demand point positions, the aim is to minimize the differences between each demand point covered by the DC. This objective function aims to minimize the difference between DPs covered by each DC.

4.2 Simulated Annealing and Genetic Algorithm Based KNN

The experiment aimed to compare the optimal solutions for DC selection using the hybrid SA-KNN (Algorithm 2) and GA-KNN (Algorithm 3). In Algorithm 2, the fitness function is used in both the SA and GA. Initially, the population's data size is initialized. Subsequently, a maximum of 10,000 iterations has been set for validation. The DC location is selected, and its value is then utilized in the KNN algorithm for classification.

Algorithm 2. Pseudo Code of Simulated Annealing based KNN

Input: *Coordinate DC, Coordinate DP*
Output: *Highest Fit*
Steps:

```
    Start and end temperature TZero, TFinal
    The parameter λ is calculated using the method proposed by Swift et al. (2004):
```

$$\lambda = \frac{exp((log(T_{Final}) - log(T_{Zero}))}{j}$$

```
    Create C by Algorithm 1
    Let t = TZero,
    Let Fit = F(C)
    For i = 1 to ITER
        Let Fnew = F(Cnew)
        Diff = fnew - fit, P = exp (Diff/t)
        If Fnew ≥ fit
            Fit = Fnew, C = Cnew
        Else
            If random value (0,1) ≤ P
                Fit = Fnew, C = Cnew
            End If
        End If
        t = lamda × t
    End For
```

Algorithm 3. Pseudo Code of Genetic Algorithm based KNN

```
Input:  Coordinate DC, Coordinate DP
   Output: Highest Fit
   Steps:
Start and end temperature TZero, TFinal
Define lamda (equation 2)
 Create C = Algorithm 1
 Let t = TZero,
 Let Fit = F(C)
 For i = 1 to ITER
    Let Fnew = F(Cnew)
   Diff = fnew - fit, P = exp(Diff/t)
   If Fnew ≥ fit
       Fit = Fnew, C = Cnew
   Else
     If random value (0,1) ≤ P
       Fit = Fnew, C = Cnew
     End If
   End If
   t = lamda × t
 End For
```

The parameter λ can be calculated using Eq. 3 [41].

$$\lambda = \frac{exp((log(T_{Final}) - log(T_{Zero}))}{j} \tag{3}$$

where:

```
Input: Coordinate DC
       Coordinate DP
Random initialise population DP
Determine fitness of population (Algorithm 1)
For Highest Fit
       Select parents from population.
       Crossover and generate new population.
       Perform mutation on new population.
       Calculate fitness (algorithm 1) for new population.
End For
Output: Highest Fit
```

5 Result and Analyst

The hybrid SA-KNN and GA-KNN have been evaluated using five datasets created with DP populations of values 100, 200, 300, 400, and 500, as displayed in Table 1. The implementation was carried out using a Python-based simulation in Spyder. The results were obtained through simulation parameters adjusted using Microsoft Excel as a simple database for storing data and the Anaconda software.

In Table 2, the results using KNN are presented, with K defined as 3. The table displays the overlap of the DP when DP is 100 and 400, indicating a waste of resources in distributing food. Furthermore, when DP is defined as 200 and 400, there are 20 left-out instances, which means that during a disaster, these areas are not covered.

In Fig. 3, one example is shown for DP 400. The green triangle represents DC(k), and the areas covered by DC are shown in light blue, yellow, and pink. The white color indicates the uncovered area.

The solution involves incorporating the left-out areas (indicated by white color) through optimization. We utilize hybrid SA-KNN and GA-KNN to determine the ‘best’ solution during flood occurrences, applicable in disaster relief. Distribution during disasters, such as providing food or items, impacts costs, including time and the involvement of volunteers. Figure 4 illustrates the population and associated costs. The graph demonstrates significant differences, particularly when the population size is 300.

Table 3 shows the results from the fitness function for hybrid SA and hybrid GA. The highest value indicates that it takes more resources but has a lower fitness function, showing the best output. However, if the DP is 500, the results show that Hybrid SA is better than Hybrid GA.

Table 1 Summarize the parameter and values

Parameter	Value
Number of Population (DP)	100, 200, 300, 400, 500
Distribution Centre (DC)	DC3
Deployment area	5000*3000 m

Table 2 Result obtained using KNN

DP	DC (k)	Total DP covered	Left out	Overlap
100	3	100	0	*20*
200	3	180	*20*	0
300	3	300	0	0
400	3	400	0	*20*
500	3	480	20	0

Fig. 3 Displays DP 400 with the DC shown, and the white circle indicates the DP left uncovered

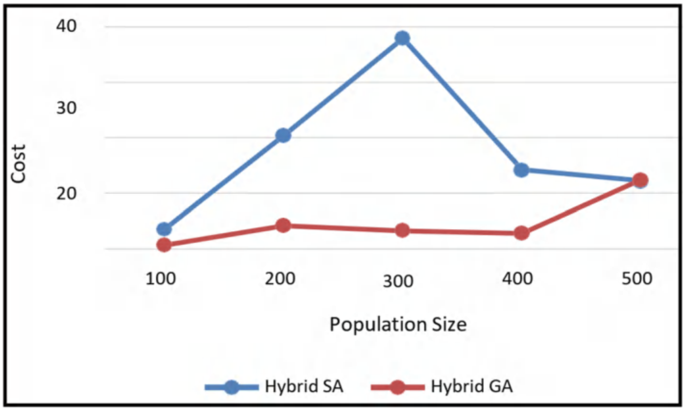


Fig. 4 Fitness value for hybrid SA and hybrid GA

Table 3 Result from fitness function for hybrid SA and hybrid GA algorithms

DP	DC (k)	Hybrid SA	Hybrid GA	Differences
100	3	3.56	0.78	2.78
200	3	20.44	4.23	16.21
300	3	38.00	3.33	34.67
400	3	14.22	2.89	11.33
500	3	12.22	12.44	−0.22
Total		88.44	23.67	

Fig. 5 Result of hybrid SA experiment with 300 DP

In Fig. 5, all the DPs are covered, and there are no overlaps between each DC. The purple circles in the population map represent DP1. The distribution of yellow circles is dominated by DP2, while the blue circles represent DP3. Although it covers all the DPs, the distribution for each DC is not balanced, which may affect the logistic cost in real implementation.

Figure 6 shows that the Hybrid GA produced better coverage, and it is mostly equal. The coordinates for DP1, DP2, and DP3 are the same in both Figs. 5 and 6, with the distribution of purple in DC1, light blue in DC2, and dark blue in DC3. All DPs are covered with optimal positions of DCs selected. Figure 6 demonstrates a more balanced distribution compared to Fig. 5. The analysis presented in this section indicates that the hybrid GA performs better than the hybrid SA. GA is effective in escaping local optima. Utilizing the hybrid GA can assist in distributing items during a disaster.

Fig. 6 Result of hybrid GA experiment with 300 DP

6 Conclusion

This research successfully identified the most effective optimization algorithm for achieving equal distribution during natural disasters. It aids in pinpointing optimal distribution center locations to ensure nearly equal distribution of DPs. Experiments with the KNN algorithm revealed issues with overlap during distribution. Enhancing the KNN algorithm with SA and GA optimizations showed that the hybrid GA approach outperforms the hybrid SA approach in ensuring equal distribution without overlaps. Cost is a crucial factor during disasters, and the study's main goal is to minimize expenses. A limitation of this research is the reliance on predefined DCs; future work should focus on automating the determination of DCs and specifying the number of volunteers involved.

References

1. Sani D D/iya, Muhd Barzani Gasim, Mohd Ekhwan Toriman, and Musa G Abdullahi, 'Floods in Malaysia : historical reviews, causes, effects and mitigations approach', *Int. J. Interdiciplinary Res. Innov.*, vol. 2, no. 4, pp. 59–65, 2014.
2. T. Mai *et al.*, 'Defining flood risk management strategies: A systems approach', *Int. J. Disaster Risk Reduct.*, vol. 47, p. 101550, 2020, https://doi.org/10.1016/j.ijdrr.2020.101550.
3. N. W. Chan, 'Resilience and recovery in Asian disasters', *Resil. Recover. Asian Disasters*, no. August, 2015, https://doi.org/10.1007/978-4-431-55022-8.
4. H. Abualkhair, E. J. Lodree, and L. B. Davis, 'Managing volunteer convergence at disaster relief centers', *Int. J. Prod. Econ.*, vol. 220, no. February, 2020, https://doi.org/10.1016/j.ijpe.2019.05.018.

5. H. H. Chiang, C. W. Ting, E. Chao, and K. J. Chen, 'Using tabletop exercises to evaluate nurses' clinical performance of hazardous materials disaster management: A cross-sectional study', *Nurse Educ. Today*, vol. 87, no. April, 2020, https://doi.org/10.1016/j.nedt.2020.104358.
6. S. Ahmad *et al.*, 'Proposed Framework for the Flood Disaster Management Cycle in Malaysia', *Sustain.*, vol. 14, pp. 1–21, 2022.
7. Anon, 'Kerajaan perlu beri tumpuan selesaikan punca banjir - Aktivis alam sekitar | Astro Awani. (2022)', [Online]. Available: https://www.astroawani.com/berita-malaysia/kerajaan-perlu-beri-tumpuan-selesaikan-punca-banjir-aktivis-alam-sekitar-398746
8. M. Bin Mokhtar and M. F. Ahmed, 'Managing a road as a river to mitigate the impact of urban flash floods', *J. Flood Risk Manag.*, vol. 15, no. 4, pp. 2–3, 2022, https://doi.org/10.1111/jfr3.12849.
9. H. S. Rosmadi, M. F. Ahmed, M. Bin Mokhtar, and C. K. Lim, 'Reviewing challenges of flood risk management in Malaysia', *Water (Switzerland)*, vol. 15, no. 13, pp. 1–21, 2023, https://doi.org/10.3390/w15132390.
10. T. R. Bhuiyan, A. C. Er, N. Muhamad, and J. J. Pereira, 'The socioeconomic impact of climate-related hazards: flash flood impact assessment in Kuala Lumpur, Malaysia', *Nat. Hazards*, vol. 109, no. 2, pp. 1509–1538, 2021, https://doi.org/10.1007/s11069-021-04887-3.
11. L. Lee, 'Malaysian manufacturers fear heavy losses from floods', *Reuters*, 2021, [Online]. Available: https://www.reuters.com/markets/commodities/malaysian-manufacturers-fear-heavy-losses-floods-2021-12-20/
12. A. Aiman, 'Flood losses "Could amount to RM20 billion"', *FMT Media Sdn Bhd*, 2021, [Online]. Available: https://www.freemalaysiatoday.com/category/nation/2021/12/24/flood-losses-could-amount-to-rm20-billion/
13. M. S. Abdullah, 'Siap siaga hadapi banjir. Utusan Malaysia', *02 November, p. 8*, 2016.
14. R. DAVIES, 'Philippines—8 killed in floods and landslides, thousands evacuate', 2020, [Online]. Available: http://floodlist.com/asia/philippines-floods-caraga-cagayan-december-2020
15. F. N. I. ASIA, 'Malaysia—hundreds displaced by floods in Sabah', 2020, [Online]. Available: http://floodlist.com/asia/malaysia-floods-sabah-october-2020
16. R. Davies, 'AFRICA EUROPE OCEANIA Malaysia—Thousands evacuate floods in Kelantan , Terengganu and Pahang', no. August, 2020.
17. P. Cunningham and S. J. Delany, 'K-nearest neighbour classifiers', *Mult Classif Syst*, vol. 54, no. 6, 2007, https://doi.org/10.1145/3459665.
18. Y. J. Zheng, S. Y. Chen, and H. F. Ling, 'Evolutionary optimization for disaster relief operations: A survey', *Appl. Soft Comput.*, vol. 27, pp. 553–566, 2015, https://doi.org/10.1016/j.asoc.2014.09.041.
19. R. Z. Farahani, M. SteadieSeifi, and N. Asgari, 'Multiple criteria facility location problems: A survey', *Appl. Math. Model.*, vol. 34, no. 7, pp. 1689–1709, 2010, https://doi.org/10.1016/j.apm.2009.10.005.
20. G. Goyal, 'Impact and analysis of improved bilateral filter on TEM images', *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4, pp. 5121–5124, 2014, [Online]. Available: https://www.ijsr.net/archive/v3i6/MDIwMTQ0NTQ=.pdf
21. X. G. Li and X. Wei, 'An improved genetic algorithm-simulated annealing hybrid algorithm for the optimization of multiple reservoirs', *Water Resour. Manag.*, vol. 22, no. 8, pp. 1031–1049, 2008, https://doi.org/10.1007/s11269-007-9209-5.
22. Y. Osa, 'Growing role of NGOs in disaster relief and humanitarian assistance in East Asia', *A Grow. Force Civ. Sociery's Role Asian Reg. Secur.*, pp. 66–89, 2013, [Online]. Available: https://goo.gl/d8vYWp
23. 23 R. Maharjan and S. Hanaoka, 'Warehouse location determination for humanitarian relief distribution in Nepal', *Transp. Res. Procedia*, vol. 25, pp. 1151–1163, 2017, https://doi.org/10.1016/j.trpro.2017.05.128.
24. G. Fetter *et al.*, 'An optimization model for humanitarian relief volunteer management An optimization model for humanitarian relief volunteer management Pamplin College of Business Pamplin College of Business', *Proc. 6th Int. ISCRAM Conf.—Gothenburg, Sweden*, no. August 2016, 2009.

25. E. Fix and J. L. Hodges, 'Discriminatory analysis. Nonparametric discrimination: Consistency properties (Report)', *USAF Sch. Aviat. Med. Randolph Field, Texas*, 1951, https://doi.org/10.2307/1403797.
26. W. Ismail, I. A. A. Q. Al-Hadi, C. Grosan, and R. Hendradi, 'Improving patient rehabilitation performance in exercise games using collaborative filtering approach', *PeerJ Comput. Sci.*, vol. 7, pp. 1–29, 2021, https://doi.org/10.7717/PEERJ-CS.599.
27. M. Fereiduni and K. Shahanaghi, 'A robust optimization model for distribution and evacuation in the disaster response phase', *J. Ind. Eng. Int.*, vol. 13, no. 1, pp. 117–141, 2017, https://doi.org/10.1007/s40092-016-0173-7.
28. F. Liu *et al.*, 'Evolution of heuristics: towards efficient automatic algorithm design using large language model', 2024, [Online]. Available: http://arxiv.org/abs/2401.02051
29. W. Ismail, R. Hassan, and S. Swift, 'Detecting Leukaemia (AML) Blood cells using cellular automata and heuristic search', *Cohen P.R., Adams N.M., Berthold M.R. Adv. Intell. Data Anal. IX. IDA 2010. Lect. Notes Comput. Sci.*, vol. 6065, pp. 54–66, 2010, https://doi.org/10.1007/978-3-642-13062-5.
30. W. Ismail, 'Automatic detection and classification of apple-a survey', *Brunel Univ. London. Thesising*, vol. 52, no. 2, 2012, https://doi.org/10.21647/icctest/2017/49057.
31. F. Majzoubi, L. Bai, and S. S. Heragu, 'The EMS vehicle patient transportation problem during a demand surge', *J. Glob. Optim.*, vol. 79, no. 4, pp. 989–1006, 2021, https://doi.org/10.1007/s10898-020-00965-1.
32. Z. Michalewicz and M. Schoenauer, 'Evolutionary algorithms for constraint parameter optimization problems', *Evol. Comput.*, vol. 4, pp. 1–32, 1996. https://doi.org/10.1162/evco.1996.4.1.1.
33. H. Jia, F. Ordóñez, and M. M. Dessouky, 'Solution approaches for facility location of medical supplies for large-scale emergencies', *Comput. Ind. Eng.*, vol. 52, no. 2, pp. 257–276, 2007, https://doi.org/10.1016/j.cie.2006.12.007.
34. G. H. Tzeng and Y. W. Chen, 'The optimal location of airport fire stations: A fuzzy multi-objective programming and revised genetic algorithm approach', *Transp. Plan. Technol.*, vol. 23, no. 1, pp. 37–55, 1999, https://doi.org/10.1080/03081069908717638.
35. L. Yang, B. F. Jones, and S. H. Yang, 'A fuzzy multi-objective programming for optimization of fire station locations through genetic algorithms', *Eur. J. Oper. Res.*, vol. 181, no. 2, pp. 903–915, 2007, https://doi.org/10.1016/j.ejor.2006.07.003.
36. G. Chai, C. Fang, X. Gao, and Q. Zhao, 'A cost-based study on highway traffic emergency rescue sites location using heuristic genetic algorithm', *J. Comput. Inf. Syst.*, vol. 7, no. 2, pp. 507–514, 2011.
37. C. F. Han and C. Zhang, 'Genetic algorithm for solving problems in emergency management', *5th Int. Conf. Nat. Comput. ICNC 2009*, vol. 4, no. October, pp. 259–264, 2009, https://doi.org/10.1109/ICNC.2009.333.
38. A. Khan *et al.*, 'Color image segmentation using genetic algorithm with aggregation-based clustering validity index (CVI)', *SIViP*, vol. 13, no. 5, pp. 833–841, 2019, https://doi.org/10.1007/s11760-019-01419-2.
39. D. Verma, V. P. Vishwakarma, and S. Dalal, 'A hybrid self-constrained genetic algorithm (Hsga) for digital image denoising based on psnr improvement', *Adv. Bioinformatics, Multimedia, Electron. Circuits Signals*, vol. 1064, pp. 135–153, 2020, https://doi.org/10.1007/978-981-15-0339-9_12.
40. A. Hiassat, A. Diabat, and I. Rahwan, 'A genetic algorithm approach for location-inventory-routing problem with perishable products', *J. Manuf. Syst.*, vol. 42, pp. 93–103, 2017, https://doi.org/10.1016/j.jmsy.2016.10.004.
41. S. Swift, A. Tucker, V. Vinciotti, N. Martin, C. Orengo, X. Liu and P. Kellam, 'Consensus clustering and functional interpretation of gene-expression data'. *Genome Biol.*, vol. 5, pp. 1–16, 2004. https://doi.org/10.1186/gb-2004-5-11-r94.

Evaluating Advanced Dimensionality Reduction Techniques for Effective Clustering of High-Dimensional Adolescent Development Data

Khoula Said Al. Abri and Manjit Singh Sidhu

Abstract This study compared several dimensionality reduction methods to analyze complex, high-dimensional data in adolescent development. Initially, we use different encoding methods, such as binary, ordinal, and one-hot encoding methods, to ensure the dataset's consistency. Then, to make our dataset less complicated, we used kernel PCA, t-Distributed Stochastic Neighbor Embedding (t-SNE), Uniform Manifold Approximation and Projection (UMAP), Locally Linear Embedding (LLE), and Isometric Mapping (Isomap). We used the K-means clustering algorithm to find important patterns and then used different validation metrics, such as the Silhouette Score, Davies-Bouldin Score, and Calinski-Harabasz Score, to compare how well different dimensionality reduction techniques worked. Our results showed that that Isomap and UMAP consistently show the best performance across all metrics, with Isomap leading slightly. For instance, Isomap's SS is 11.5% higher than UMAP's, and it also outperforms UMAP in DBS and CHS by 17.1 and 4.3%, respectively. Kernel PCA performs decently but falls short, with Isomap surpassing it by 27.0% in compactness and 67.2% in cluster separation. LLE and t-SNE are the weakest, with Isomap's SS being 11.9% higher than LLE's and 22.5% higher than t-SNE's, and its CHS significantly outperforming both, by 188.2 and 73.0%, respectively. This study provides valuable insights into adolescent development, offering a clearer understanding of patterns and relationships, and suggesting that Isomap is particularly effective in capturing hidden features within the data.

Keywords Dimensionality reduction · Adolescent development · Encoding techniques · K-Means clustering · Clustering · Validation metrics

K. S. Al. Abri (✉) · M. S. Sidhu
Department of Graphics and Multimedia, College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Kajang Selangor, Malaysia
e-mail: kobaid@su.edu.om

M. S. Sidhu
e-mail: manjit@uniten.edu.my

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_2

1 Introduction

Personality in adolescence undergoes significant changes, influenced by emotional, cognitive, and social developments [1]. In this period, personalities change to the greatest possible extent. During the teenage stage of life, developmental theories can guide parents and educators by helping their counselors more accurately interpret or help shape parenting relationships, peer influence, schooling aspirations, cultural backgrounds, etc [2]. According to developmental theories, the environment plays a significant role in adolescent decision-making and maladaptive adulthood [3]. Adolescents can have choices for the development of these attributes as well, such as morality and objective career goals. As such, these theories can explain how gender differences, growing up, and socioeconomic status all play a role in determining the specific trajectory and velocity of adolescent development. Understanding these factors enables carers to guide adolescents into healthier trajectories and help them meet a range of needs [4].

Three developmental theories provide insightful perspectives on the development of young people: Donald Super's Developmental Self-Concept Theory suggests that adolescents actively explore possible careers through various opportunities before making life commitments between ages 15–24 [5]. According to the Social Cognitive Theory of Albert Bandura, one's career interests and decisions are largely influenced by an interactive environment and levels of self-enhancement belief [6]. According to John Holland's Theory of Job Choice, teenagers tend to be happy and motivated to choose employment that fits with their inherent attributes. This theory classifies people into six distinct personality types [7]. These theories together suggest that a teenager's ability to make decisions regarding their future employment will be influenced by the interplay between their basic self-awareness, the effect of surrounding conditions, and their unique skills [8].

There is a notable gap in the ability to interpret high-dimensional datasets related to adolescent development, as recent studies show that simple statistical tests and regression techniques often struggle to work with nonlinear datasets [9]. For example, a study by Rajula et al. [10] and Stachl et al. [11] demonstrated these conventional methods may produce inaccurate or misleading results leading to oversimplified conclusions [10] and [11]. In addition, Miraftabzadeh et al. [12] highlighted how K-means clustering efficiently groups similar data points, even in nonlinear datasets related to consumer behavior [12].

In this study, we integrate five advanced dimensionality reduction techniques: Kernel PCA (KPCA), t-SNE, t-SNE clustered, UMAP, LLE, Isomap with K-Means clustering. The objective of the paper is to identify the most efficient strategy for conducting exploratory analysis on a nonlinear and heterogeneous dataset of adolescent personalities. It assists in capturing of the relationships within the data set, hence helping in representing the complex relationships within the data set. Therefore, it would be possible to offer a concrete and extensive theoretical foundation concerning adolescent development. In this way, we make a list of these techniques for the further comparative analysis and outline the further integrated strategy that allows overcoming the existing differences of traditional approaches. This work puts the basis and the new standards in high dimensional personality data analysis.

KPCA is an extension of the typical PCA that utilizes kernel functions to find the non-linear relations among data points. Additionally, this is very helpful for image processing and bioinformatics [13]. However, unlike PCA, KPCA is computationally time-consuming, especially when working with large datasets, since one should compute and store the kernel matrix, which leads to poor scalability. When you have a non-linear dimensional reduction problem, KPCA works well. For example, [14] showed that KPCA has improved interpretability, leading to a better understanding of complex data.

t-SNE was designed to visualize high-dimensional data sources in the fewest dimensions possible, usually 2 or 3, where the local structure is preserved. It is widely used in applications such as genomics and image recognition. However, methods such as t-SNE are computationally expensive and subject to the choice of hyperparameters, such as perplexity. In addition, it is not easy to interpret the global structure of the data with t-SNE plots. However, this local structure in high-dimensional data can be beneficial if you genuinely want to explore your features. Still, the interdependency of its performance with initialization and hyperparameters also makes it hard to benchmark fairly [15].

UMAP, however, offers a balanced approach by effectively preserving both local and global geometries of the underlying data with high computational efficiency. Nowadays, this has become a standard procedure to reduce dimensions and waste space, hence keeping the data ready for modelling in machine learning [16]. Although much faster than t-SNE, UMAP also needs hyperparameter tuning to get the best results. If the hyperparameters aren't optimized properly, the interpretation of the pre-trained embedding can be hard to determine. UMAP has been proven to be effective for clustering pipelines, but only when merged with an algorithm like K-Means [17].

LLE is designed to retain local data structures, making it suitable for exploring manifold structures where maintaining neighborhood relationships is crucial, such as in sensor network data analysis. However, LLE is sensitive to noise and does not scale well with large datasets due to its approach to restoring neighborhood structures [18]. Consequently, for datasets with significant local structure but also sensitivity to noise, LLE can produce very noisy results. Enhancements proposed by [18] have improved its robustness in noisy data environments.

On the other hand, Isomap is adept at preserving global geometric relationships, making it suitable for investigating non-linear complexities within datasets, such as those found in image processing and pattern recognition. Isomap requires computing geodesic distances between all point pairs, resulting in significant computational costs [19]. This complexity limits its practicality for large datasets. Despite this, Isomap effectively captures the underlying structure of the data. The efficiency improvements proposed by [20] offer a more practical solution for larger datasets.

The combination of clustering algorithms, such as K-Means, and dimensionality reduction methods can be used to achieve better data segmentation quality [21]. Lower dimensionality helps K-Means form clusters more easily and meaningfully.

K-Means cluster data points into groups by minimizing in-cluster variance, making it very effective and clear for observation results after dimensional reduction [22].

In this work, we explored career interest, career maturity, and decision-making responses using contemporary dimensionality reduction and clustering algorithms. In the initial stage of our approach, we thoroughly prepared our dataset. This involved using binary and ordinal encoding, as well as one-hot encoding. We aimed to ensure that the dataset was free from any inconsistencies and properly prepared for further analysis. We used Kernel PCA, t-SNE, UMAP, LLE, and Isomap techniques for dimensionality reduction in conjunction with K-Means clustering to create the mentioned clusters. We measured performance using the SS, DBS, and CHS. Better results will be shown lower in each plot. This provides the basis to use these features as predictors and learn more from our data. The clear cluster traction on ISOMAP reflects its effectiveness for this type of analysis. The current work provides significant implications for further research on adolescent career-linked qualities and decision-making skills.

2 Methodology

2.1 Data Preprocessing and Transformation

One of our goals was to get the dataset ready for analysis. For this, we followed some preprocessing steps which helped in making the dataset clean and consistent and ready for analysis. This cleanup is necessary before applying Dimensionality Reduction and Clustering on the data efficiently. Our dataset contains 1,004 samples and 141 features.

First, we loaded the dataset into Pandas DataFrame. Next, to handle missing values, we replaced empty strings with NaN and removed rows where all entries were NaN, ensuring our dataset did not have any gaps that could disrupt the analysis. Then, we renamed each column to a standard format like Q1, Q2 to maintain consistency and ease of reference, and converted the columns to the ’category’ data type for more efficient processing.

Furthermore, we used different encoding techniques because our dataset has diff for the career interest questions (Q1–Q54), we simplified the answers by using binary encoding, changing ‘Yes’ to 1 and ‘No’ to 0, making the responses easier to analyze as numerical values.

$$\text{binary}_{\text{encoded}_{\text{value}}} = \begin{cases} 1 \ if \ \ response = Yes \\ 0 \ if \ \ response = No \end{cases} \tag{1}$$

We encoded the responses for career maturity (Q55–Q99) and decision-making (Q100–Q136) using an ordinal scale from 1 to 5. This scale reflects the natural order of the responses.

$$A = \begin{cases} 1 \; if \;\; response = stronglydisagree \\ 2 \; if \;\; response = \quad disagree \\ 3 \; if \;\; response = \quad notsure \\ 4 \; if \;\; response = \quad agree \\ 5 \; if \;\; response = \; stronglyagree \end{cases} \tag{2}$$

For categorical features like region, we used one-hot encoding. This creates a new binary column for each category. For example, one-hot encoding converts the region column into seventh columns from (region _A to region_SJ).

$$one\_hot\_encoded\_value_{i,j} = \begin{cases} 1 \; if \;\; data\, po\mathrm{int}\, i \; belongs \; to \; category\, j \\ 0 \qquad\qquad otherwise \end{cases} \tag{3}$$

We converted range responses, such as weight and height, to their numerical midpoints. For example, '60–70 kg' became 65 kg.

$$\text{midpoint} = \frac{(lower\_bound + pper\_bound)}{2} \tag{4}$$

2.2 Advanced Dimensionality Reduction Techniques and Clustering

In this section, we first employed dimensionality reduction techniques such as Isomap, t-SNE, UMAP, KPCA, and LLE because the nature of our dataset because our dataset is highly nonlinear and heterogeneous, as indicated by the poor performance of linear models and the complex relationships observed during preliminary analyses, making these advanced methods essential for uncovering meaningful patterns and enhancing clustering performance. We further tuned several parameters of these techniques to find out which one would work best to enhance our K-means clustering performance. The dimensionality reduction and clustering process is then validated with validation metrics, which include SS, DBS, and CHS. How each of the dimensionality reduction techniques and its parameters can result in better clustering results and accurate elicitation of insight is explained in the following sections.

KPCA improves principal component analysis with the use of kernel methods to analyze data in a higher-dimensional space. In contrast to regular PCA, which works only with plain linear relationships, KPCA uses special functions called kernels to map data into a space where more complex patterns can be detected. We tested several kernel functions to find the one that best described the structure of our data [23].

1. Linear Kernel: Projects data into a higher-dimensional space while maintaining linear relationships:

$$k\left(x_i, x_j\right) = x_i^T x_j \tag{5}$$

2. Polynomial Kernel: Captures polynomial interactions between features:

$$k\left(x_i, x_j\right)= \left(x_i^T x_j + c\right)^d \tag{6}$$

where c adjusts the linear term's influence, and d controls polynomial complexity. Higher d fits more complex interactions, while higher ccc provides added flexibility but may introduce noise.

3. The Radial Basis Function (RBF) Kernel captures complex, nonlinear relationships with:

$$\left(x_i, x_j\right) = exp\left(\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{7}$$

where σ regulates the kernel's smoothness. Smaller σ emphasizes local structures, while larger σ captures global trends. After selecting the appropriate kernel, we compute the kernel matrix:

$$k_{ij} = k\left(x_i, x_j\right) \tag{8}$$

to capture similarities in the higher-dimensional space. We then centered this matrix to stabilize the analysis and solved the eigenvalue problem:

$$\tilde{k}\alpha = \lambda\alpha \tag{9}$$

where α are the eigenvectors and λ are the eigenvalues. The eigenvectors that correspond to the largest eigenvalues define the principal components in the transformed space

t-SNE aims to keep similar responses close together in the lower-dimensional space, making it easier to identify patterns and relationships. Perplexity is a key parameter in t-SNE that influences how many neighboring responses are considered when calculating similarities. It essentially controls the balance between local detail and global structure:

- Low perplexity (e.g., 5) emphasizes very local structures, making it easier to see fine distinctions between individual student responses.
- High perplexity (e.g., 50) includes more neighbors in the similarity calculation, helping to reveal broader patterns and general relationships among responses.

For our dataset, using a perplexity value between 5 and 50 effectively balances detailed individual patterns with overall trends among students. t-SNE reduces the dimensions of the data by minimizing the difference between two types of similarity measures: one based on distances in the high-dimensional space and another based on distances in the lower-dimensional space [24].

1. High-dimensional similarities: These calculations are derived from the distances between data points within the original, high-dimensional distance. The formula for this is:

$$P_{ij} = \frac{\exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})}{\sum_{k\neq 1} \exp(-\frac{\|x_i - x_j\|^2}{2\sigma^2})} \tag{10}$$

   where, $\|x_i - x_j\|$ represents the Euclidean distance between data points $x_i$and x_j and σ (perplexity parameter) adjusts the scale of these distances. This formula calculates the probability P_{ij} that x_i and x_j are neighbors in the high-dimensional space.

2. Low-dimensional similarities: After reducing the dimensions, t-SNE calculates similarities in this new space. The formula for these is:

$$Q_{ij} = \frac{(1 + \|y_i - y_j\|^2)^{-1}}{\sum_{k\neq 1} (1 + \|y_i - y_j\|^2)^{-1}} \tag{11}$$

   where $\|y_i - y_j\|$ denotes the Euclidean distance separating the locations $y_i$andy_j in the lower-dimensional space. This formula gives the probability Q_{ij} that y_i and y_j

3. Kullback-Leibler (KL) Divergence: The difference between the high-dimensional and low-dimensional similarities is measured using the KL divergence, which t-SNE aims to minimize.

$$C = \sum_{k\neq 1} P_{ij} \log \frac{P_{ij}}{Q_{ij}} \tag{12}$$

   where, C represents the overall cost function that measures how well the low-dimensional representation preserves the high-dimensional similarities.

   In dimensionality reduction techniques like UMAP, LLE, and Isomap, the choice of neighborhood size is crucial for preserving the data's intrinsic structure. Neighborhood sizes of 5, 10, and 30 are often chosen to balance between capturing local details and global patterns in the data.**UMAP** reduces dimensions while preserving **local and global data structures** by constructing a high-dimensional graph and optimizing it to a lower-dimensional graph [20]. The optimization objective is:

$$\operatorname{argmin} \sum_{(i,j)\in E} w_{ij} \mathrm{d}(\mathrm{x}_i, \mathrm{x_j}) \tag{13}$$

where w_{ij} are edge weights in the graph, and $d(x_i, x_j)$ is the distance between points x_i and x_j in the low-dimensional space.

LLE preserves local neighborhood relationships by reconstructing each data point linearly from its neighbors [25]. The reconstruction error minimized is:

$$\in_i = \|x_i - \sum_j W_{ij}X_j\|^2 \tag{14}$$

where, W_{ij} are the reconstruction weights. Solving the eigenvalue problem:

$$(I - W)^T(I - W)_y = \lambda y \tag{15}$$

provides the low-dimensional embeddings, where I is the identity matrix, W is the weight matrix, and λ are the eigenvalues.

Isomap captures the global geometric structure by computing geodesic distances and performing Multidimensional Scaling (MDS). The geodesic distance matrix D is calculated, and the classical MDS is performed by centering the matrix [26].

$$k = -\frac{1}{2}HDH \tag{16}$$

where, H is the centering matrix. Solving the eigenvalue problem:

$$K\alpha = \lambda\alpha \tag{17}$$

where, K is the centered distance matrix, α are the eigenvectors, and λ are the eigenvalues. The eigenvectors corresponding to the largest eigenvalues provide the low-dimensional embeddings.

After applying dimensionality reduction, we use K-Means clustering to identify the most important features within each cluster [27]. To identify the best number of clusters, we evaluate multiple clustering validation metrics across a range of cluster counts (2–10).

1. Silhouette Score $s(i)$:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \tag{18}$$

where a_i is the mean distance from the data point i to other points within its cluster, and b_i is the mean distance from i to points in the closest neighboring cluster [27]. A higher s_i means better clustering because it shows that points are closer to their own cluster (a_i) and further from other clusters (b_i).

2. Davies-Bouldin Score (DB):

$$DB = \frac{1}{\mathrm{k}}\sum_{i=1}^{k} max_{j\neq i}\left(\frac{s_i + s_j}{d_{ij}}\right) \tag{19}$$

where, s_i represents the mean distance from data point i to its cluster's center, while d_{ij} denotes the distance between the centroids of clusters i and j. A lower DDS reflects superior clustering, indicating that the clusters are more compact and clearly distinct [29].

3. Calinski-Harabasz Score (CH):

$$CH = \frac{tr(B)}{tr(W)} \times \frac{N - K}{K - 1} \tag{20}$$

where, $tr(B)$ stands for the trace of the matrix that captures the dispersion between clusters, $tr(W)$ is the trace of the matrix that captures within-cluster dispersion, N denotes the total number of data points, and K is the number of clusters [29]. An increased CH score signifies better clustering, as it demonstrates that the clusters are more clearly separated and internally compact.

3 Result

3.1 Results and Insights from Data Processing

This section presents the result of different encoding techniques. Binary encoding was applied to the Career Interest columns (Q1–Q54). Each 'Yes' response was encoded as 1 and 'No' as 0. Table 1 shows the encoding results for the first five student's responses in the dataset.

Ordinal encoding was applied to the Career Maturity columns (Q55–sQ99) and Decision-Making columns (Q100–Q136) with a scale from 1 (strongly disagree) to 5 (strongly agree). Tables 2 and 3 illustrate the encoding results for the first five individuals.

One-hot encoding was applied to the Demographic Features (Q137–Q142), resulting in binary columns representing each category. Table 4 displays encoding

Table 1 Binary encoding for career interest columns (Q1–Q54)

Student no	Q1	Q2	Q3	Q4	Q5	...	Q50	Q51	Q52	Q53	Q54
0	1	1	0	0	1	...	1	0	0	0	1
1	1	0	1	1	1	...	1	0	0	1	1
2	0	1	1	0	0	...	0	1	1	1	1
3	0	1	1	0	0	...	1	0	0	1	1
4	0	0	0	1	0	...	1	1	0	1	1

Table 2 Ordinal encoding for career maturity columns (Q55–Q99)

Student no	Q55	Q56	Q57	Q58	Q59	...	Q95	Q96	Q97	Q98	Q99
0	4.0	5.0	3.0	3.0	5.0	...	5.0	5.0	4.0	3.0	3.0
1	3.0	5.0	1.0	5.0	5.0	...	5.0	4.0	3.0	3.0	5.0
2	2.0	4.0	1.0	1.0	5.0	...	4.0	4.0	3.0	2.0	2.0
3	4.0	4.0	5.0	3.0	5.0	...	5.0	5.0	5.0	5.0	5.0
4	5.0	5.0	5.0	4.0	3.0	...	3.0	5.0	2.0	3.0	3.0

Table 3 Ordinal encoding for decision-making columns (Q100–Q136)

Student No	Q100	Q101	Q102	Q103	Q104	...	Q132	Q133	Q134	Q135	Q136
0	5.0	5.0	5.0	4.0	4.0	...	5.0	5.0	3.0	5.0	5.0
1	1.0	4.0	3.0	5.0	1.0	...	1.0	4.0	4.0	5.0	4.0
2	2.0	4.0	5.0	2.0	2.0	...	5.0	4.0	1.0	3.0	3.0
3	5.0	5.0	5.0	5.0	5.0	...	5.0	5.0	5.0	5.0	5.0
4	4.0	4.0	4.0	3.0	2.0	...	3.0	3.0	5.0	5.0	5.0

result for (Q137–Q142) and Table 5 represent one hot encoding result for different regions.

Table 4 Encoding for demographic features (Q137–Q142)

Student no	Q137	Q139	Q140		Q141	Q142
0	0	1	85		55.0	152.5
1	1	1	85		85.0	177.5
2	1	1	85		75.0	177.5
3	1	1	85		75.0	172.5
4	0	1	75		55.0	162.5

Table 5 One-hot encoding results for region

Student no	R_A	R_B	R_C	R_D	R_E	R_F	R_G	R_H	R_I	R_J
0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	1	0
2	0	0	0	0	0	0	0	0	1	0
3	0	0	0	0	0	0	0	0	1	0
4	0	0	1	0	0	0	0	0	0	0

3.2 Advanced Clustering Performance Evaluation Using Dimensionality Reduction Methods

In this study, we compared the performance of KPCA, t-SNE, UMAP, LLE, and Isomap based on their clustering performance using three key metrics: SS, DBS, and CHS. We avoid biased conclusions and make sure that the dimensionality reduction techniques are evaluated fairly by combining all three metrics along with the visual evidence from the 3D clustering visualizations.

The results of Isomap with n_neighbors = 5, presented in Table 6, show that with 3 clusters, it achieved optimal results, yielding a SS of 0.6475 and a DBS of 0.4375, indicating excellent cluster separation and cohesion. However, the best CHS of 4060.0051 was observed with 6 clusters, suggesting that this configuration offers a superior balance of clustering compactness and separation. These metrics underscore the effectiveness of Isomap with this parameter setting in capturing complex patterns in the dataset.

Despite these high scores, in Fig. 1, Isomap's clustering shows a clear separation between clusters, with some areas of overlap, particularly between the brown and light blue clusters. The data points within each cluster are well-separated, indicating strong compactness. The overlap, however, suggests that the clusters, while distinct, exhibit a smooth transition between certain groups. The brown cluster, for example, has a dense core that naturally extends into the space of the light blue cluster, reflecting a nuanced and continuous relationship rather than strict boundaries.

As detailed in Table 7, the UMAP method, with n_neighbors = 10, resulted in effective clustering with three clusters. The performance metrics include an SS of 0.5825, a DBS of 0.5288, and a CHS of 3031.9881, showcasing robust cluster separation and compactness.

UMAP's SS of 0.5825 is lower than Isomap's SS for three clusters, which indicates that Isomap achieves better overall separation between clusters with a difference of approximately 11.5%. UMAP's DBS of 0.5288 is higher than Isomap's DBS for three clusters, reflecting that Isomap has more compact and well-defined clusters, with a difference of about 17.1%. Isomap's CHS of 3163.8315 is also higher than UMAP's CHS for three clusters, suggesting that Isomap provides better cluster

Table 6 Isomap clustering performance (n_neighbors = 5)

No. clusters	SS	DBS	CFS
2	0.5683	0.6218	1535.2567
3	0.6495	0.4383	3163.8315
4	0.5755	0.5425	3309.4019
5	0.5752	0.5541	3927.3666
6	0.5576	0.5989	4204.4379
7	0.5284	0.6605	4111.9649
8	0.5097	0.6833	4065.5224
9	0.4920	0.7116	4009.1937

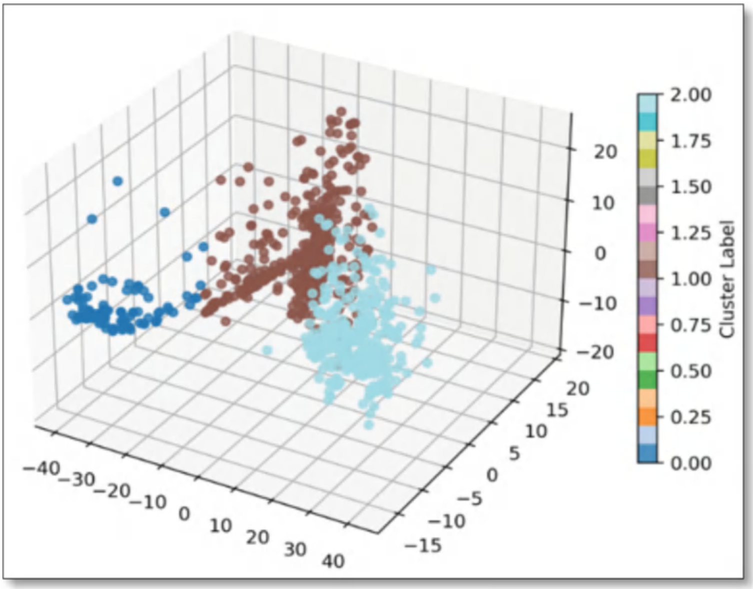


Fig. 1 Isomap 3D clustering visualization (n_neighbors: 5, n_clusters: 3)

Table 7 UMAP clustering performance (n_neighbors = 10)

No. clusters	SS	DBS	CHS
2	0.5808	0.5742	2219.3519
3	0.5825	0.5288	3031.9881
4	0.4937	0.7251	2827.3330
5	0.4390	0.8474	2832.2647
6	0.4306	0.8319	2688.1659
7	0.4200	0.7956	2699.4492
8	0.4105	0.8231	2677.4182
9	0.3954	0.8396	2666.8526
10	0.3947	0.8052	2635.2484

compactness and separation with a difference of approximately 4.3%, which indicates that Isomap achieves better overall separation between clusters with a difference of approximately.

Figure 2 demonstrates UMAP's excellent cluster separation with three distinct clusters. Within each cluster, the data points are tightly grouped, showcasing good compactness. When compared with Isomap, UMAP's clustering shows slightly better overall balance in terms of cluster compactness and separation, making them also a good choice for detailed data analysis.

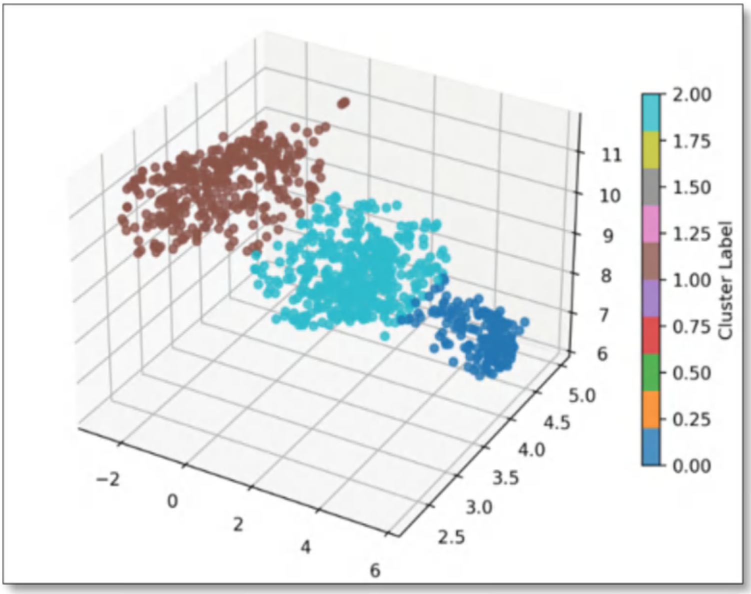


Fig. 2 UMAP 3D clustering (n_neighbors: 10, n_clusters: 3)

The KPCA analysis, as outlined in Table 8, reveals that using a polynomial kernel led to optimal clustering with two clusters. This setup achieved an SS of 0.6420 and a DBS of 0.6007, showing that the clusters are distinct and cohesive. Furthermore, a CHS of 1892.2047 was noted with two clusters, suggesting a good balance between cluster compactness and separation.

KPCA showed a high SS, which is better than UMAP by approximately 10.2%, indicating excellent cluster separation. However, its DBS was higher than both UMAP and Isomap by about 14 and 37%, respectively, indicating more overlap. Its

Table 8 KPCA clustering performance (Polynomial Kernel)

No. clusters	SS	DBS	CHS
2	0.6420	0.6007	1892.2047
3	0.5370	0.6738	1991.2848
4	0.4621	0.7313	2074.5616
5	0.4289	0.8259	1969.7689
6	0.4257	0.8520	1842.2439
7	0.4204	0.8546	1796.1147
8	0.4089	0.8551	1795.3614
9	0.3831	0.8342	1796.9735
10	0.3878	0.8230	1813.5953

CHS was significantly lower than both UMAP and Isomap by approximately 37.6 and 36.5%, respectively.

The clusters in Fig. 3 not only overlap, but they also show that the gray cluster has few data points, which shows that KPCA is not as good as UMAP and Isomap at making compact and well-separated clusters. This spread-out arrangement means the data points are far apart, which makes it harder for KPCA to create clear cluster boundaries.

The results of LLE with n_neighbors = 10, presented in Table 9, show that with 4 clusters, it achieved optimal results, yielding an SS of 0.5804 and a DBS of 0.5322, indicating good cluster separation and cohesion. However, the highest CHS of 1933.0029 was observed with 10 clusters, suggesting that this configuration offers a favorable balance of clustering compactness and separation.

In Fig. 4, the LLE method was used with n_neighbors = 10 and 4 clusters. The figure shows distinct clusters colored in different shades, but there is noticeable overlap between them. Specifically, the pink, black, and orange clusters show significant intermingling, indicating poor separation. The data points within each cluster are somewhat dispersed, suggesting moderate compactness. This overlap highlights that while LLE can preserve neighborhood structures to some extent, it does not achieve clear and well-separated clusters as effectively as other methods like UMAP and Isomap. This makes LLE less suitable for tasks requiring distinct cluster boundaries.

In Table 10, the results of the t-SNE analysis show that a perplexity setting of 50 yielded the best clustering results with two clusters. This configuration achieved

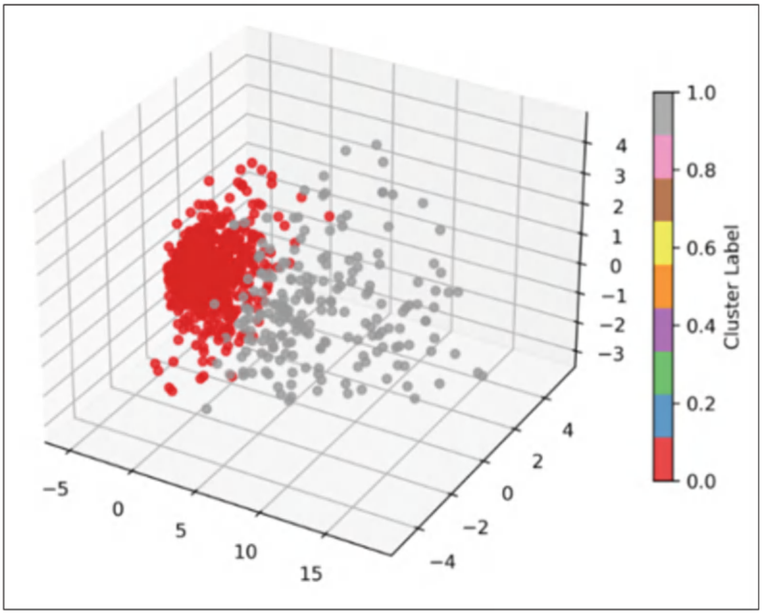


Fig. 3 KPCA 3D clustering visualization (Polynomial kernel)

Table 9 LLE clustering performance (n_neighbors = 10)

No. clusters	SS	DBS	CHS
2	0.5163	0.8870	653.2349
3	0.5760	0.7135	928.9520
4	0.5804	0.5322	1097.8864
5	0.5610	0.5567	1380.6936
6	0.5298	0.5531	1635.7460
7	0.5082	0.5960	1789.6391
8	0.5015	0.6500	1825.6479
9	0.4870	0.6501	1895.3247
10	0.4691	0.6805	1933.0029

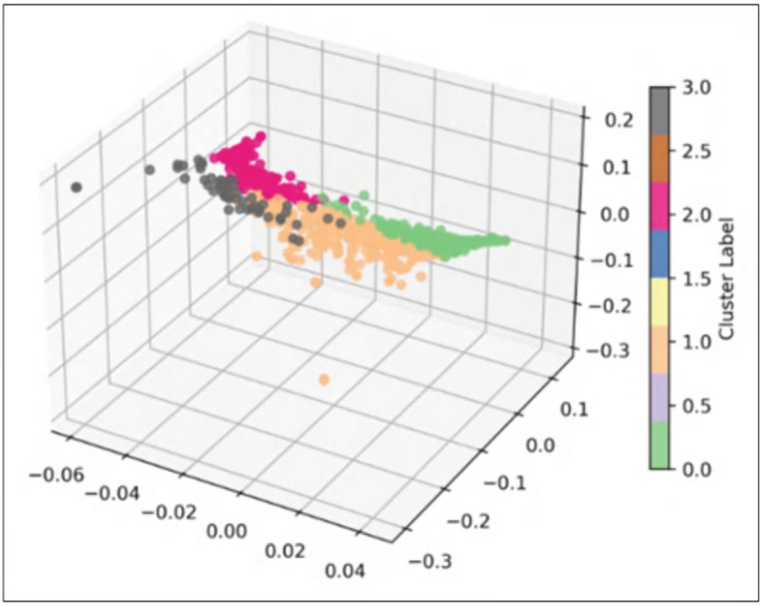


Fig. 4 LLE 3D clustering (n_neighbors: 10, n_clusters: 4)

an SS of 0.5302, a DBS of 0.6652, and a CHS of 1828.8448, reflecting reasonable cluster separation, although not as strong as other methods.

In Fig. 5, the t-SNE method was used with a perplexity setting of 50 and 2 clusters. The figure shows two distinct clusters, colored red and blue. However, there is a significant overlap between the clusters, indicating poor separation. The data points within each cluster are tightly packed, showing good compactness within the clusters but failing to maintain distinct boundaries between them. This overlap suggests that t-SNE, while useful for visualizing local structures, is less effective for achieving

Table 10 t-SNE clustering performance (Perplexity 50)

No. clusters	SS	DBS	CHS
2	0.5302	0.6652	1828.8448
3	0.4676	0.7584	1810.3261
4	0.4524	0.8604	1586.1056
5	0.4088	0.8430	1514.4378
6	0.3835	0.8794	1523.8186
7	0.3633	0.8687	1469.6270
8	0.3521	0.8835	1427.4162
9	0.3476	0.8894	1391.2219
10	0.3498	0.8649	1377.8534

clear and distinct cluster separation. We conclude that t-SNE is better suitability for exploratory visualization than precise clustering analysis.

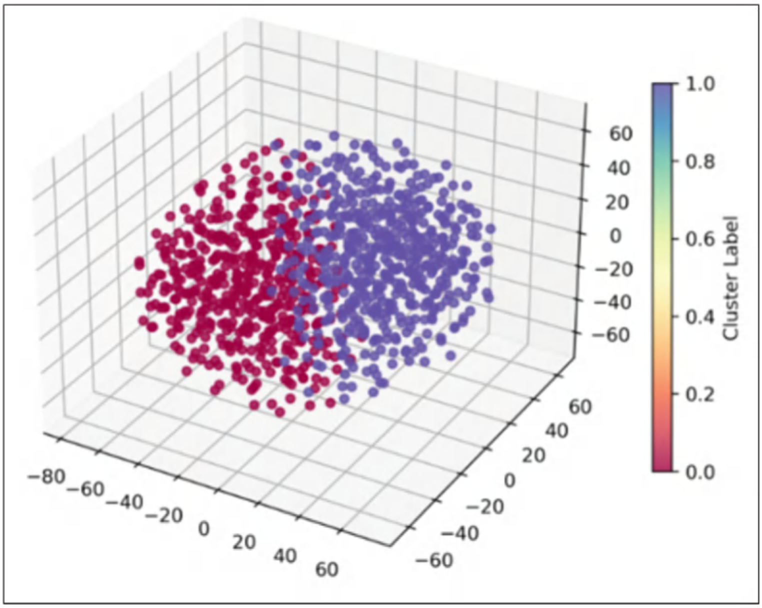


Fig. 5 t-SNE 3D clustering (Perplexity: 50, n_clusters: 2)

3.3 *Isomap Feature Extraction and Ranking of Key Differentiating Factors*

Using Isomap, we extracted and ranked the most informative features from the dataset based on average differences in centroid values between clusters. The elbow method was used to determine the optimal number of features, which is 82, by plotting explained variance against the number of features to find the "elbow point," where the variance change rate shifts sharply.

the feature ranking is categorizing features by their level of importance in differentiating between clusters for example Highly important features (average differences > 1.3) include confidence in career decisions, ensuring accurate professional information, the impact of conflicting advice, and the ability to find alternatives and solutions. Moderately important features (average differences between 1.0 and 1.3) involve planning and analyzing issues, pessimism about future career decisions, collecting and analyzing information about future professions, awareness of university conditions, and identifying and analyzing problems. Less important features (average differences < 1.0) encompass knowledge about job market needs and self-awareness, specific tasks and interests, external influences and personal preferences, and enjoyment of specific activities.

4 Discussion

Our findings strongly recommend combining advanced dimensionality reduction techniques with clustering algorithms when analyzing high-dimensional data related to adolescent development. We took a critical look at KPCA, t-SNE, UMAP, LLE, and Isomap in terms of their effectiveness for these types of tasks. We discovered that the underlying methods work well with complex patterns in the data.

Isomap also demonstrated an excellent result across validation metrics, with a Silhouette Score of 0.6495, a Davies-Bouldin Score of 0.4383, and a Calinski-Harabasz Score of 3163.8315 with good visualization among the other techniques. Robustness was excellent, with the clusters showing satisfactory separation and compactness. This makes the application of Isomap very useful in revealing hidden relationships in adolescents' data. UMAP also did a fantastic job preserving global geometric relationships, so this is another excellent choice for researching detailed data overviews. These tools are powerful instruments for researchers and practitioners to enhance their understanding of the many facets of adolescent development.

Consequently, our findings have broad implications; approaches like Isomap and UMAP would help educators and even psychologists gain valuable insight, and they might even help policymakers understand what really influences the development of adolescents. This insight would allow them to equip teenagers with life skills and help them make beneficial life decisions. From a practical perspective, educators and psychologists could tailor their advice and scaffolding activities to more personalized and effective strategies designed for adolescents. These techniques, in fact, could lead to more personalized and effective educational and psychological support strategies by identifying key factors differentiating adolescents' personalities.

While we are excited with these results, the fact that the dataset used consisted of actual data obtained directly from teenagers significantly strengthens our study. This real-world dataset enhances the relevance and authenticity of our findings, ensuring that the insights derived are not just theoretical but are grounded in actual adolescent experiences. This authenticity makes our conclusions highly applicable in real-world contexts, providing valuable guidance for educators, psychologists, and policymakers.

We believe that future studies should test these techniques on a broader variety of datasets to further corroborate and build on our findings. Additionally, we observed that dimensionality reduction techniques like t-SNE or LLE can have variable performance depending on the choice of hyperparameters. This underscores the need for ongoing work on hyperparameter optimization and standardization to ensure reproducibility across different settings. Some techniques, such as KPCA and t-SNE, also proved computationally intensive, which may limit their practicality for very large datasets.

Moving forward, we believe that the next step in this line of research should be to apply these techniques to a wider array of high-dimensional datasets. It would also be beneficial to study in greater detail the effects of different preprocessing steps and hyperparameter settings, which would allow for further fine-tuning of each applied dimensionality reduction technique. Improved visualization techniques would also enhance the interpretation and communication of clustering and dimensionality reduction results, making these methods even more practical. Furthermore, exploring the practical applications of this empirical evidence in educational and psychological interventions could help develop targeted strategies that effectively support adolescent development.

5 Conclusion

Our work highlights the significant potential of advanced dimensionality reduction and clustering techniques in the context of adolescent development. Isomap stands out for its strong visualization capabilities, enabling the capture of complex patterns and relationships within the data. These findings open new possibilities for more effective analysis and intervention strategies in adolescent development, ultimately helping educators and practitioners provide teenagers with more robust support as they grow both personally and professionally.

References

1. Ran, L.-Y., et al.: Personality traits predict treatment outcome of an antidepressant in untreated adolescents with depression: An 8-week, open-label, flexible-dose study. J. Affect. Disord. 350(11), 102–109, https://doi.org/10.1016/j.jad.2024.01.015 (2024).
2. Sugiyarlin, Supriatna, M.: Adolescent's career maturity. In: International Conference on Educational Psychology and Pedagogy (ICEPP 2019), pp. 232–235, https://doi.org/10.2991/assehr.k.200130.121 (2020).
3. Ting, C., Kwee, T.: The Application of Career Theories in Teachers' Professional Development and Career Decision: A Literature Review. Universal J. Educ. Res. 8(9), 3997–4008, https://doi.org/10.13189/ujer.2020.080925 (2021).
4. Tawafak, R.M., Alfarsi, G., Romli, A., Jabbar, J., Malik, S.I., Alsideiri, A.: A review paper on student-graduate advisory expert system. In: 2020 International Conference on Computing and Information Technology, ICCIT 2020, pp. 7–11, https://doi.org/10.1109/ICCIT-144147971.2020.9213794 (2020).
5. Jena, L., Nayak, U.: Theories of Career Development: An analysis. Indian J. Nat. Sci. 10(60), 23515–23523 (2020).
6. Ilmiani, A.M., Wahdah, N., Mubarak, M.R.: The application of albert bandura's social cognitive theory: A process in learning speaking skill. Ta'lim al-'Arabiyyah: J. Pendidi-kan Bahasa Arab & Kebahasaaraban 5(2), https://doi.org/10.15575/jpba.v5i2.12945 (2021).
7. Zainudin, Z.N., Lee, W.R., Nor, A.M., Yusop, Y.M.: The Relationship of Holland Theory in Career Decision Making. J. Clin. Res., https://doi.org/10.31838/jcr.07.09.165 (2020).
8. Donnelly, K.W., Cawthon, T.W.: Introduction: Overview and Application of Career Development Theories. Wiley Online Libr. 2019(10), 9–17, https://doi.org/10.1002/ss.20303 (2019).
9. Khosravi, B., et al.: Demystifying Statistics and Machine Learning in Analysis of Structured Tabular Data. J. Arthroplasty 38(10), 1943–1947, https://doi.org/10.1016/j.arth.2023.08.045 (2023).
10. Rajula, H.S.R., Verlato, G., Manchia, M., Antonucci, N., Fanos, V.: Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment. Medicina (Lith.) 56(9), 1–10, https://doi.org/10.3390/medicina56090455 (2020).
11. Stachl, C., et al.: Personality Research and Assessment in the Era of Machine Learning. Eur. J. Pers. 34(5), 613–631, https://doi.org/10.1002/per.2257 (2020).
12. Miraftabzadeh, S.M., Colombo, C.G., Longo, M., Foiadelli, F.: K-Means and Alternative Clustering Methods in Modern Power Systems. IEEE Access 11, 119596–119633, https://doi.org/10.1109/ACCESS.2023.3327640 (2023).
13. Geetha, V., Anbumani, V., Parameshwaran, R., Gomathi, S.: Savitzky Golay and KPCA-based Optimal Discrete Wavelet Transform Architecture for Image Compression. Microprocess. Microsyst. 91, 104511, https://doi.org/10.1016/j.micpro.2022.104511 (2022).
14. Bounoua, W., Bakdi, A.: Fault detection and diagnosis of nonlinear dynamical processes through correlation dimension and fractal analysis-based dynamic kernel PCA. Chem. Eng. Sci. 229, https://doi.org/10.1016/j.ces.2020.116099 (2021).
15. Chen, X., et al.: Health diagnosis of concrete dams with continuous missing data for assessing structural deformation based on tSNE–AHC algorithm and deep transfer learning. Struct. 57, https://doi.org/10.1016/j.istruc.2023.105134 (2023).
16. Qing, X., et al.: Essential Spectral Pixels-based Improvement of UMAP Classifying Hyperspectral Imaging Data to Identify Minor Compounds in Food Matrix. Talanta 273, 125845, https://doi.org/10.1016/j.talanta.2024.125845 (2024).
17. de Zarzà, I., de Curtò, J., Calafate, C.T.: UMAP for Geospatial Data Visualization. Procedia Comput. Sci. 225, 1661–1671, https://doi.org/10.1016/j.procs.2023.10.155 (2023).
18. Wang, J., Wong, R.K.W., Lee, T.C.M.: Locally Linear Embedding with Additive Noise. Pattern Recognit. Lett. 123, 47–52, https://doi.org/10.1016/j.patrec.2019.02.030 (2019).

19. Qu, H., Li, L., Li, Z., Zheng, J.: Supervised Discriminant Isomap with Maximum Margin Graph Regularization for Dimensionality Reduction. Expert Syst. Appl. 180, 115055, https://doi.org/10.1016/j.eswa.2021.115055 (2021).
20. Hajderanj, L., Chen, D., Dudley, S., Gilloppe, G., Sivy, B.: Novel parameter-free and parametric same degree distribution-based dimensionality reduction algorithms for trustworthy data structure preserving. Inf. Sci. (N. Y.) 661, 120030, https://doi.org/10.1016/j.ins.2023.120030 (2024).
21. Jia, W., Sun, M., Lian, J., Hou, S.: Feature Dimensionality Reduction: A Review. Complex Intell. Syst. 8(3), 2663–2693, https://doi.org/10.1007/s40747-021-00637-x (2022).
22. Rouhi, A., Bouyer, A., Arasteh, B., Liu, X.: Two-pronged Feature Reduction in Spectral Clustering with Optimized Landmark Selection. Appl. Soft Comput. 161, 111775, https://doi.org/10.1016/j.asoc.2024.111775 (2024).
23. Tonin, F., Tao, Q., Patrinos, P., Suykens, J.A.K.: Deep Kernel Principal Component Analysis for Multi-level Feature Learning. Neural Networks 170, 578–595, https://doi.org/10.1016/j.neunet.2023.11.045 (2024).
24. Hajibabaee, P., Pourkamali-Anaraki, F., Hariri-Ardebili, M.A.: An empirical evaluation of the t-SNE algorithm for data visualization in structural engineering. In: Proceedings of the 20th IEEE International Conference on Machine Learning and Applications, ICMLA 2021, pp. 1674–1680, https://doi.org/10.1109/ICMLA52953.2021.00267 (2021).
25. Liu, S., Chen, Z., Jiao, F.: Detection of Maize Seed Germination Rate Based on Improved Locally Linear Embedding. Comput. Electron. Agric. 204, 107514, https://doi.org/10.1016/j.compag.2022.107514 (2023).
26. Anowar, F., Sadaoui, S., Selim, B.: Conceptual and Empirical Comparison of Dimensionality Reduction Algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE). Comput. Sci. Rev. 40, 100378, https://doi.org/10.1016/j.cosrev.2021.100378 (2021).
27. Shutaywi, M.: Silhouette Analysis for Performance Evaluation in Machine. Entropy 23(6), 1–17, https://doi.org/10.3390/e23060759 (2021).
28. 29. Ros, F., Riad, R., Guillaume, S.: PDBI: A Partitioning Davies-Bouldin Index for Clustering Evaluation. Neurocomputing 528, 178–199, https://doi.org/10.1016/j.neucom.2023.01.043 (2023).
29. 30. Yu, H., Zhou, C., Bao, J., Yao, H., Yan, P., Wang, Q.: Analysis and Effect Evaluation of Offshore Wind Power Output Characteristics Based on Gaussian Mixed Clustering. Procedia Comput. Sci. 224, 389–394, https://doi.org/10.1016/j.procs.2023.09.053 (2023).

A Decade-Long Bibliometric Analysis of Electronic Health Record Research Trends (2014–2024)

Muhammed Ibrahim, Moamin A. Mahmoud, Mohammed A. Al-Sharafi, and Abba Hassan

Abstract This study analyzes global research trends in electronic health records through a bibliometric approach, using the Scopus database, 18,555 articles published between 2014 and 2024 were retrieved, focusing on author co-authorships, publication outputs, the co-occurrence of author keywords, and affiliated countries. The findings show a steady increase in publications from 2014 to 2023. Researchers from the United States, China, and India made the most significant contributions to global publications. Affiliation institutions of the authors include Harvard Medical School, which leads with an impressive publication count is followed by Mayo Clinic which ranks second in terms of publications, while Columbia University come third in publications. In terms of co-authorship link strength, Gadekallu Thippa Reddy emerged as the author with the highest total link strength of 146. This Followed by Zhang Tao securing the second position with a total link strength of 42, other authors like Tang Buzhou, Wang Jiaqi and Xu Hui were also identified with their significance contribution in the field of EHR implementation each with their substantial link strengths of 17, 18 and 19 respectively. In terms of authors' citations, Ghayvat Hemant leads with 67 citations, followed by Zhang Tao with 42 citations, and Xie Feng with 33 citations. The study concludes by providing broader summary of electronic health record trends, using data from the Scopus database. The findings indicate that in

M. Ibrahim (✉) · A. Hassan
College of Graduate Studies, Universiti Tenaga Nasional, Selangor, Malaysia
e-mail: ibrahimcrc321@yahoo.com

Faculty of Computing, Nigerian Army University, Biu, Borno State, Nigeria

A. Hassan
e-mail: minshah@utm.my; SShamini@uniten.edu.my

M. A. Mahmoud
College of Computing and Informatics, Universiti Tenaga Nasional, Selangor, Malaysia
e-mail: moamin@uniten.edu.my

Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Selangor, Malaysia

M. A. Al-Sharafi
IRC for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_3

the past decade there is a significant growth, there is projections in the future it will continue to increase.

Keywords Electronic Health Record (EHR) · Health informatics · Blockchain · Co-authorship · Co-occurrences

1 Introduction

Governments and industries worldwide are increasingly driving the digitization of healthcare systems, a trend reflected in various initiatives across different countries and sectors [16, 20]. Recent IT advancements have transformed health records through digitalization, improving methods for collecting, processing, storing, and sharing data. This digitization enhances data portability, facilitates seamless sharing among healthcare organizations, and allows public health administrators to conduct surveys and develop policies more effectively. Patients also gain better access to their health information, enabling more effective self-management of their care [13]. In 2009, the U.S. president signed the Health Information Technology for Economic and Clinical Health (HITECH) Act, part of the American Recovery and Reinvestment Act. This legislation was designed to promote the widespread adoption of EHR, aiming to enhance benefits for both patients and society by improving healthcare efficiency and outcomes [16, 20].

EHR facilitate seamless and accurate access to patient information for healthcare providers and organizations. However, implementing EHR systems can be expensive, with costs ranging from $162,000 in primary care settings to over $1 billion in large healthcare organizations. EHRs, often known as Electronic Medical Records (EMRs), have transformed healthcare by enabling efficient information sharing and enhancing patient transparency [7]. EMRs are crucial components of e-health applications, containing comprehensive health-related data, including demographics, diagnoses, medications, laboratory results, medical images, and clinical notes [2]. They serve as legal records created within hospital environments and are a primary source of data for EHR systems [13]. Initially intended to improve the efficiency and accessibility of healthcare systems, EHRs have evolved to support various applications in clinical informatics and epidemiology [2]. EHRs encompass all essential administrative and clinical data related to patient care, including progress notes, medications, vital signs, medical history, immunization records, laboratory data, and radiology reports [13]. They are used for extracting medical concepts, clustering diseases and patients, modeling patient outcomes, predicting diseases, and supporting data-driven clinical decisions [2]. The integration of Information Technology (IT) into EHR systems has improved patient care experiences, enhanced population health, and reduced healthcare costs [13]. The potential benefits of EHR systems, such as better public healthcare management, online patient access, and efficient data sharing, have also garnered significant research interest [9, 11]. While most clinicians acknowledge the value of EHRs, they also express a strong desire for significant improvements. Experiences with EHRs have the potential to influence career plans and overall satisfaction among healthcare professionals [17]. EHRs are defined as a digital representation of an individual's medical history, containing essential administrative and clinical data pertinent to their healthcare [1]. eHealth has shown significant promise not only in alleviating strain on overwhelmed medical systems but also in influencing health-related behaviors and facilitating the long-term management of chronic diseases [22]. An EHR system requires a specialized team of healthcare professionals and IT specialists to regularly monitor, maintain, repair, and update its various programs [7]. However, these capabilities can also pose substantial challenges to maintaining the confidentiality necessary for delivering sensitive services [8]. Blockchain technology known for supporting secure transactions among network participants has been reported to be utilized in EHRs. Although commonly linked to Bitcoin, blockchain technology uses a digital ledger protocol with public key cryptography to create an unchangeable, time-stamped chain of health information. This technology provides a new foundation for healthcare systems, promoting decentralization, immutability, and interoperability [3]. The advent of healthcare 3.0 which aligns with the Web 3.0 concept, enabled the customization of how patient healthcare records were managed. User interfaces became more tailored and user-friendly resulting in optimized and personalized experiences. This era also introduced EHRs and wearable and implantable systems, allowing for real-time and widespread monitoring of patients' health [21]. The prevailing global economic recession and financial challenges encountered by numerous countries leading to reduced budgets and widespread implementation of cost-saving strategies the process of digitalization or digital transformation has persisted across various regions and nations. This ongoing digital advancement can be credited to several factors which include the growing digital requirements of individuals and their increased expectations, alongside the influence of emerging products stemming from continuously innovative technologies [12]. The benefits of EHRs become particularly apparent when they are integrated with computerized provider order entry (CPOE) systems and real-time point-of-care clinical decision support (CDS) interventions. Incorporating decision support systems into clinical care is anticipated to improve care quality and lead to substantial reductions in healthcare costs [14]. EHRs represent a valuable data source that has the potential to streamline the conduct of clinical trials. Serving as the primary repository of clinical information for patients within a healthcare system [19]. Specialized hospitals with advanced equipment are on the rise worldwide. Patients are increasingly interested in seeking care from these healthcare providers often being examined by specialized doctors for better treatment [3]. EHRs contain a vast amount of data that can be leveraged to identify potentially suitable candidates for enrolment in clinical trials. Growing evidence indicates promising opportunities for using EHR-based data acquisition [19]. Patient information which includes personal, clinical, pharmacy, and accounting data, is often stored in fragmented ways across different locations. As this data grows rapidly over time, tasks related to updating and accessing it become cumbersome, in addition to presenting security issues thus the need of EHR implementation arise [3].

Smart technologies which include the Artificial Intelligence (AI), Machine Learning (ML), Internet of Things (IoT), Augmented Reality (AR), and Virtual Reality (VR) have transformed industries such as engineering, manufacturing, automotive, computing, electronics, aerospace and defense. Healthcare systems used by providers like hospitals and general practitioners have also benefited from these technologies, becoming increasingly powerful and useful over time [21]. This bibliometric analysis will be different from the previous conducted literature review, systematic literature review, scoping review in the field of health informatics [2, 14, 18, 22] over the past decade these are the main areas researchers have discussed, there is little focus on bibliometric analysis in the field [12]. However, several critical perspectives remain inadequately addressed, significantly impacting the widespread adoption and implementation of EHR systems. These include the countries that have been the primary focus of research, the key areas of investigation within the domain, the years with the top number of publications, highest funding institutions for EHR research, the leading researchers and their affiliations. Given these gaps, a comprehensive search of the literature for bibliometric analysis was undertaken using VOSviewer and the Scopus database to extract the data. In order to drives new changes in the domain, this analysis aimed to build on previous studies and specifically explore the various viewpoints present in the literature on EHR. In order to understand the current themes their distribution and patterns of collaboration in the domain, a thorough analysis of EHR publications is important. By identifying primary contributors, their citations, institutions, and keywords commonly used, researchers may address emerging challenges through establishing collaboration networks. This study seeks to address the following research questions:

RQ1. Who are the prominent co-authors, institution affliation, countries, with their citations contributing to the study of electronic health records?

RQ2. What are the keywords frequently used in electronic health records research?

The bibliometric analysis was well-structured, starting with an introduction and methodology that explained data collection and evaluation procedures. The article then presented detailed results and discussions, analyzing information obtained through bibliometric techniques, particularly focusing on trends and findings related to electronic health records. Finally, it was concluded with the main findings in summary, an acknowledgment of its limitations, and suggestions that will guide future study.

2 Methodology

2.1 Methods

This bibliometric study aimed to explore the publication themes within electronic health record literature indexed in the Scopus database. Bibliometric analysis, a systematic and methodical approach, leverages data from academic literature databases to offer a thorough understanding of global research trends in a specific field [15]. While review articles concentrate on recent advancements, challenges, and future directions within a specific area, with the goal of identifying performance bibliometric analysis employs a quantitative approach to assess, present, and evaluate data in research studies [4, 5].

2.2 Data Search Strategy and Procedure for Selection

Data extraction was conducted on January 19, 2024, utilizing the Scopus database to focus on study publications that contain the term "Electronic Health Record" in both the title and abstract. The query used to search is as follows: TITLE-ABS-KEY ("Electronic Health Record" OR "Health Information Technology" OR "Health Information Systems" OR "Clinical Information Systems" OR "EHR Implementation" OR "EHR Adoption" OR "EHR Integration" OR "EHR Interoperability" OR "Health Data Exchange" OR "EHR Benefits and Challenges" OR "Meaningful Use of EHR" OR "EHR Security and Privacy" OR "EHR User Satisfaction" OR "EHR Usability" OR "Health Informatics" OR "Electronic Medical Records" OR "Personal Health Records" OR "EHR Standards" OR "Health Data Analytics" OR "Population Health Management" OR "Clinical Decision Support Systems" OR "Health Information Exchange" OR "EHR and Patient Engagement" OR "EHR and Quality of Care" OR "EHR Workflow"). This search produced 111,721 documents. The PRISMA 2020 guideline was utilized to refine and extract the most relevant and important studies. The study is aimed to evaluate a decade of EHR literature from 2014 to 2024. Thus, filters were applied to restrict publication within the said range. To further narrow the literature for more precise results, the following filters were added to limit only publications in English within the fields of computer science, decision sciences, and social sciences, published between 2014 and 2024. With these filters, 93,166 papers were deemed irrelevant and removed from the analysis, as shown in Fig. 1, this process led to the selection of 18,555 articles for the study.

3 Results and Discussion

A dataset which comprises a total of 18,555 articles that consists of authors citation information, bibliographic details, and author keywords was exported from Microsoft Excel (2016) to VOSviewer (version 1.6.20). VOSviewer is a robust software tool specifically developed specifically for the purpose of generating and visualizing bibliometric maps [6]. VOSviewer was used for this bibliometric analysis because it is technically designed to handle map visualization and analyses that contain large amounts of bibliometric data network [23]. Especially, because the software offers tools for creating clear visualizations of co-author networks, keyword co-occurrence

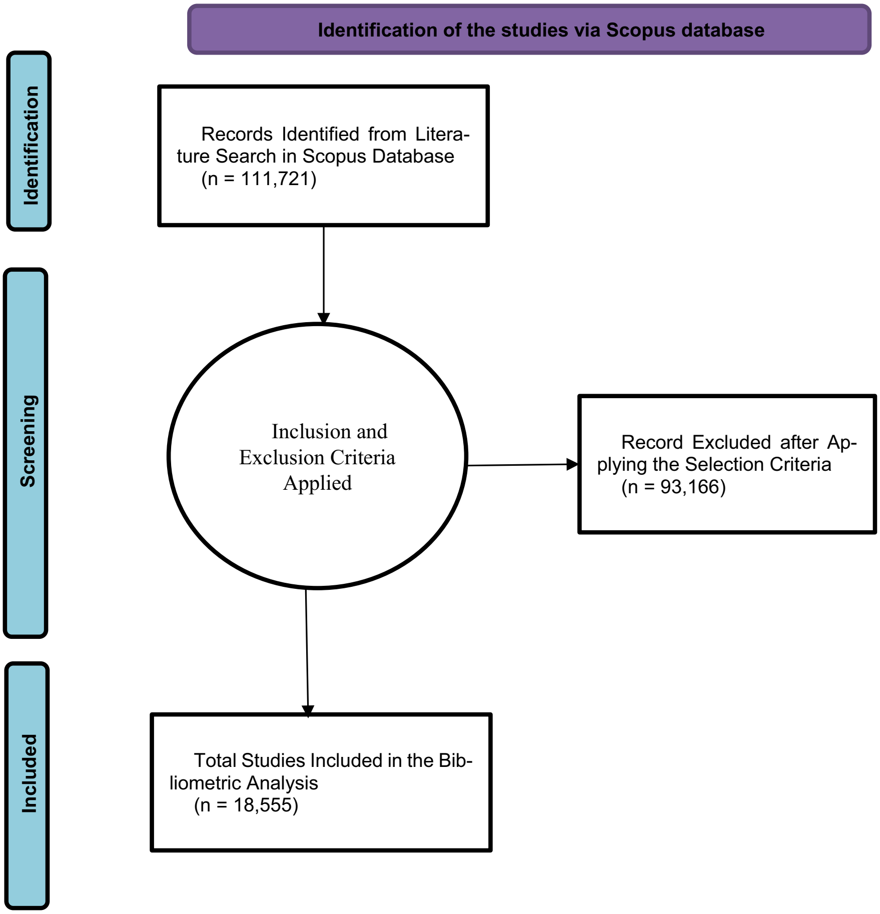


Fig. 1 Search strategy for studies and selection procedure

and citation patterns, all these would be important measures to analyze EHR bibliometric feeds [24]. Additionally, these features made VOSviewer an excellent choice for shedding light upon this EHR investigation. VOSviewer has easy-to-use interface; compatibility with Scopus data and powerful network analysis capabilities that allowed a thorough study of collaboration patterns, research topics and the most influential authors in EHR. Among its features are density visualization and clustering algorithms that allowed a better understanding of structure and evolution of EHR research between the years 2014–2024 [6].

This study's bibliometric analysis is comprehensive, examining co-authorship, institutional affiliations, author citation, and author keywords. Each of these elements is interconnected with their relationships quantified by positive numerical values by which higher values indicate stronger connections. The analysis of co-authorship, the link strengths among authors represent publications number co-authored by researchers, the total link strength indicates the extent of an author's collaborative co-authorship network. The strength of the links between an author's keywords in co-occurrence analysis reflects the frequency with which two keywords appear together in a study.

3.1 Studies Publications Per-Year

Results analysis of publication regarding EHR from 2014 to 2024 provide interesting knowledge into the evolving research field as shown in Fig. 2. In 2014, a total of 1,132 studies on EHR were published, which constituted about 6.10% of the entire publication which total to 18,555 studies. This shows a significant presence of EHR-related research in the early years of the analyzed period. This relatively modest figure suggests a moderate level of interest or engagement in EHR research during the early stages of the analyzed period. However, over time there has been a noticeable increase in the number of publications which shows a growing focus on EHR related topics in academic and scientific communities. The year 2016 was particularly significant with 1,347 studies published, accounting for around 7.26% of the total publications. This rise in research output suggests an increasing recognition of the significance of EHR systems in healthcare delivery and management. Subsequent years also experienced growth, with a significant increase in 2022 and 2023. There is increase in the number of published articles in the year 2022–2,535, representing approximately 13.66% of the total publication. In 2023, this figure increased further to 2,639 studies, constituting 14.22% of the total publications. This trend signifies the expanding interest and importance of EHR systems in healthcare research and practice.

The increase in research output in 2022 and 2023 likely indicates a heightened focus on EHR related studies that was influenced by factors such as technological advancements, policy changes, and the widespread implementation of EHR systems

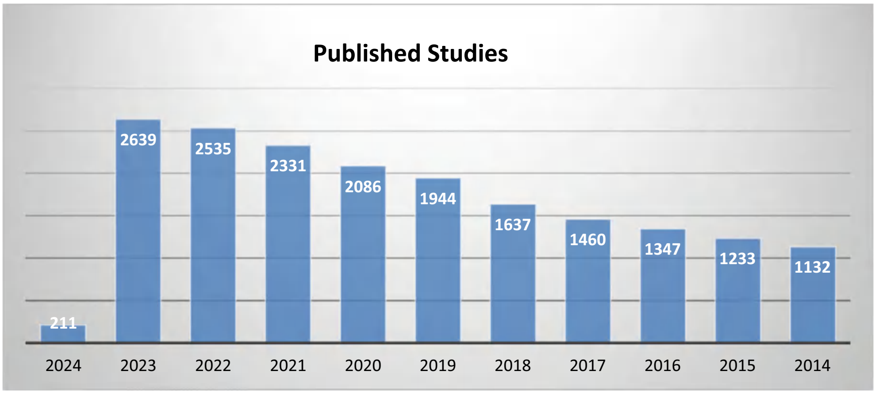


Fig. 2 Annual electronic health record publications indexed in scopus from 2014 to 2024

in healthcare settings. Thus far, this address RQ1 by showing that there has been a steady interest and contribution in the number of EHR publication over the decade globally. However, in 2024, there was a slight decline in the number of publications, with 211 studies representing 1.14% of the total. This decrease could be due to the time the search was conducted at the beginning of the year, but there is every tendency that research output in the domain will continue to increase.

3.2 Top 10 Countries Contributing to Electronic Health Record

The analysis of country-wise publication distribution in electronic health records (EHR) research from 2014 to 2024 provides a comprehensive overview of global engagement and contributions in this field. Leading the charge is the United States, with a notable 6,606 published studies comprising approximately 35.60% of the total publications. This substantial presence shows that the United States' status as a frontrunner in EHR research which might likely be driven by its advanced healthcare infrastructure, robust academic institutions, and significant investment in healthcare technology. The dominance of the United States in EHR literature suggests a strong commitment to advancing knowledge and innovation in this critical area of healthcare. Following closely behind is China, contributing 2,027 published studies, accounting for approximately 10.92% of the total publications. China's emergence as a prominent player in EHR research reflects its rapid economic development, growing healthcare sector, and strategic emphasis on technological advancement. With a large population and expanding healthcare needs China's investment in EHR research shows its commitment in making use of technology to improve healthcare delivery and outcomes for its citizens. India also emerges as a significant contributor, with 1,911 published studies, representing approximately 10.30% of the total publications. India's substantial presence in EHR literature reflects its status as a major player in the global healthcare industry, driven by a burgeoning IT sector, a vast pool of skilled professionals, and a focus on leveraging technology to address healthcare challenges. India's contributions to EHR research highlight its potential to influence the development and adoption of EHR systems not only domestically but also on a global scale.

Countries like the United Kingdom, Australia, Canada, Germany, Italy, Spain, and South Korea contribute significantly to EHR literature, showcasing diverse research landscapes and healthcare systems. These contributions reflect global interest and investment in advancing EHR technologies to improve healthcare delivery, patient outcomes, and population health. The findings emphasize the collaborative and diverse nature of EHR research, with global contributions enriching discussions and driving innovation in healthcare practices. The percentage of contribution highlights each country's proportional impact and influence in shaping scientific studies globally. It also reveals that developing nations particularly Africa, and other Asian region

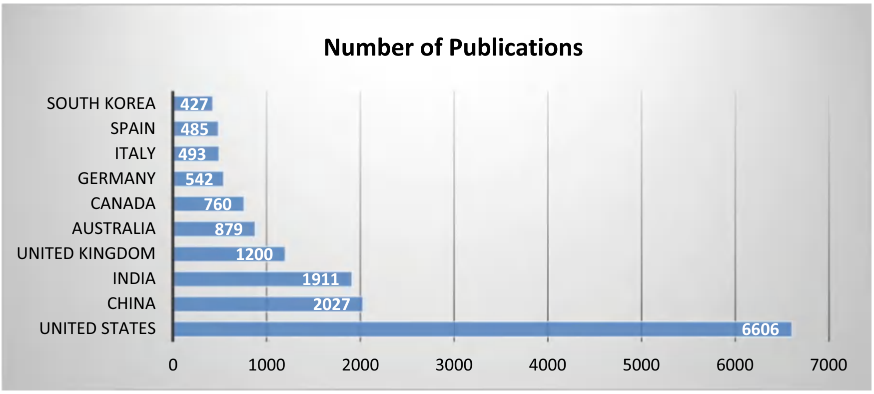


Fig. 3 Top 10 countries leading in electronic health record publications

there exist a notable knowledge gap. Lastly, these findings address RQ1, as shown in Fig. 3, which clearly illustrates countries with dominant and lesser contributions.

3.3 Authors Affiliation Distribution

The analysis of the EHR implementation demonstrates the significant impact of the top 10 institutions involved in this important area of research (see Fig. 4). Together these institutions have produced a substantial body of literature which total to 1,762 papers, that represents a significant 9.48% of all research on EHR adoption and implementation. This shows the focused efforts and specialized knowledge these institutions have dedicated to advancing understanding in this important field. At the forefront is Harvard Medical School, which leads with an impressive publication count of 302 papers, making up 1.57% of the total publications. This indicated the Harvard's central role in shaping scholarly discussions on EHR adoption, emphasizing its considerable influence within the scientific community. Close behind is Mayo Clinic, ranking second with 200 publications, equivalent to 1.03% of the overall publications. Similarly, Columbia University is a significant contributor with 194 publications, representing 1.02% of the published research.

Other institutions such as Vanderbilt University Medical Center, VA Medical Center, University of Pennsylvania, The University of Utah, Brigham and Women's Hospital, University of Washington, and University of California, San Francisco, also make substantial contributions to EHR adoption research. Together their efforts contribute to 5.67% of the total research output, highlighting the collaborative endeavors among these esteemed institutions to address the challenges and opportunities in EHR adoption. Each institution brings its unique expertise and viewpoint to the forefront, enriching discussions and fostering innovation in this evolving field.

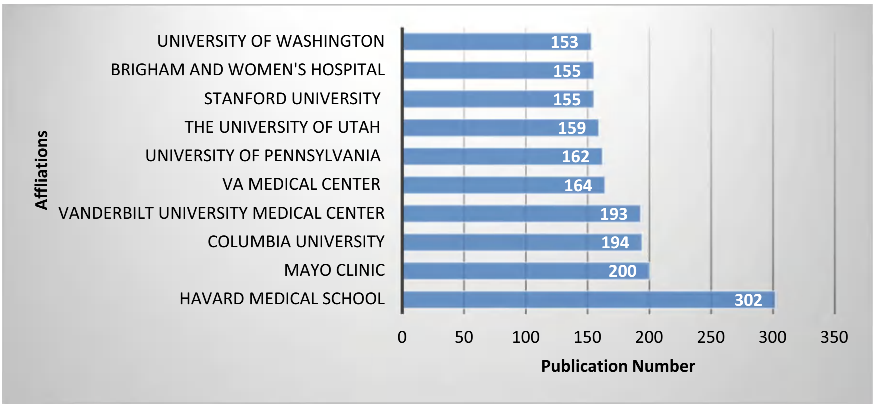


Fig. 4 Top 10 institutions by electronic health record publications indexed in scopus 2014–2024

The distribution of research output across these top institutions indicated the interdisciplinary nature of EHR adoption studies, encompassing elements of medicine, healthcare management, technology, and policy.

3.4 Co-Authorship

The analysis of co-authorship revealed that 15,057 authors published studies on EHR within the period from 2014 to 2024. Certain criteria were applied to narrow down the focus which include a maximum of 10 authors per document, a minimum of 3 documents per author, and a minimum of 5 citations. This filtering process reduced the number of authors to 157 as shown in Fig. 5. Among these authors the study identified those with the highest link strength showing their significant impact in the field. Gadekallu Thippa Reddy emerged as the author with the highest total link strength of 146. This shows that his work likely examines various critical aspects of EHR implementation. His research may have played a significant role towards influencing the adoption of EHR systems across various healthcare settings for the purpose of enhancing patient care and operational efficiency. Followed by Zhang Tao securing the second position with 42 link strength in total which shows the importance of his studies with a strong influence in the field of EHR implementation. Tang Buzhou, Wang Jiaqi and Xu Hui were also identified with their significance contribution in the field of EHR implementation each with their substantial link strengths of 17, 18 and 19 respectively, showing their significant contributions to EHR implementation research. Other authors like Li Jingsong with 12 link strength, Zhou Tianshu with11 link strength, Wang Yasha with also link strength of 9, Wu Qing has 6 link strength in total, it shows significant contribution by these authors in the field of medical informatics towards providing stable and sustainable healthcare as shown in Fig. 6.

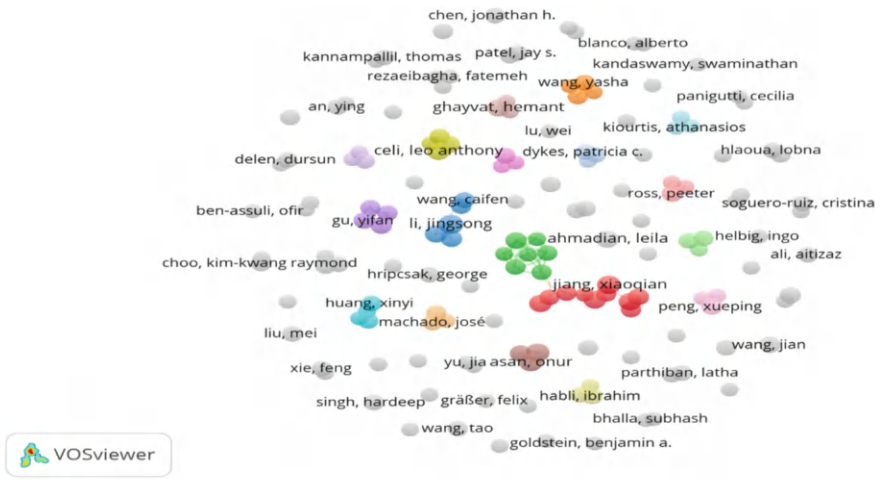


Fig. 5 Co-authorship and their link

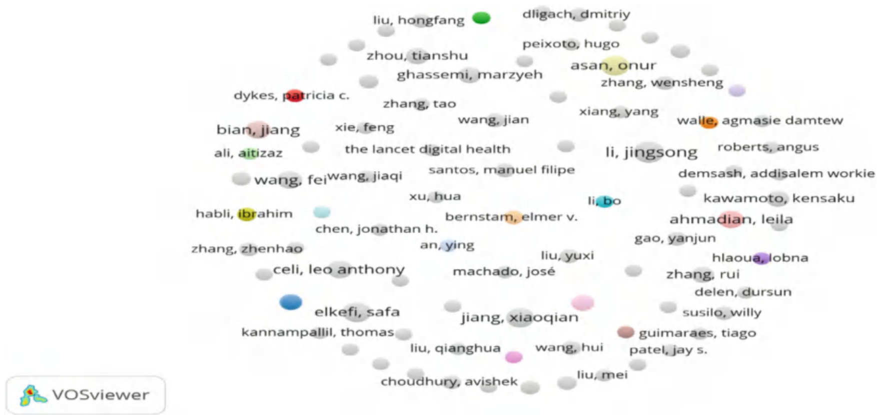


Fig. 6 Authors citation of electronic health record related publications

3.5 Authors Citation

The author citation analysis set criteria limiting to 10 authors per document, requiring at least 4 documents and 3 citations per author. From 15,057 authors, 96 met these conditions, as shown in Fig. 6. Hemant Ghayvat appeared to be with the highest citations 67 in 6 documents, and 1 link strength, emphasizing his important role in health informatics, especially in EHR implementation. Zhang Tao follows with 4 documents and 42 citations, though his link strength is 0, indicating his impact despite limited collaboration. Li Jing-song, with 9 documents—the most among the authors—has 15 citations but no link strength, highlighting his considerable contribution through sheer volume. Xie Feng (4 documents, 33 citations) and Zhang Rui (6 documents, 27 citations) also show strong individual contributions but lack network connections.

Other notable authors include Luo Xiao (4 documents, 13 citations), Ross Peteer (5 documents, 14 citations, total link strength of 1), and Habli Ibrahim, who, with 5 documents and 23 citations, shows a strong network presence with a total link strength of 4. Additional contributors such as Santos Manuel Filipe (5 documents, 3 citations), Yang Xuebing (4 documents, 7 citations), Zhi Degui (4 documents, 4 citations, total link strength of 1), Zhou Tianshu (6 documents, 7 citations), Ahmadian Leila (7 documents, 11 citations), and Marzyeh Ghassemi (6 documents, 17 citations) reflect the diverse and valuable contributions to health informatics. Despite varying citation counts and link strengths, their collective work significantly advances the healthcare sector, underscoring the importance of both individual research efforts and broader collaboration in this rapidly evolving field.

4 Keyword Co-Occurrence

A thorough co-occurrence analysis was performed using VOSviewer with data extracted from Excel, resulting in a total of 8,824 author keywords. To refine this list and identify the most impactful keywords, a criterion was set requiring each keyword to appear at least 10 times. This process narrowed the selection to 221 pertinent keywords, as illustrated in Fig. 7. This meticulous approach ensures that the keywords are both numerous and significant, offering a concentrated set for in-depth analysis. These keywords related to electronic health record research provide a comprehensive overview of the diverse areas covered in academic discussions. The analysis of keyword frequency and cumulative link strength highlights key themes in EHR implementation and health informatics. “Electronic health records” stands out with the highest link strength of 774, followed by “Machine learning” at 736 and “blockchain” at 594, indicating their strong influence in health informatics research. Blockchain, in particular, holds promise for enhancing the security and privacy of EHR systems by ensuring data accuracy and confidentiality. Keywords like “artificial intelligence (AI)” and “deep learning,” with link strengths of 482 and 449, respectively, are increasingly used in healthcare for tasks such as image analysis, predictive analytics, and natural language processing, offering new opportunities to enhance patient care and streamline operations. “Healthcare informatics,” with a link strength of 224, underscores the critical role of EHRs and informatics in modern healthcare, highlighting the need for effective health data management to support clinical decision-making and improve patient outcomes. “Telehealth” and “telemedicine” also play important roles, with link strengths of 73 and 124, reflecting the shift toward remote healthcare delivery, fueled by technological advancements and the growing availability of telehealth services. Additionally, “Access control,” with a link strength of 96, emphasizes the importance of regulating access to sensitive health information within EHR systems. Effective access control measures are crucial for maintaining data privacy and security in healthcare settings, ensuring that sensitive patient information is protected [10]. Big data and big data analytics represent keywords with total link strengths of 107 and 31 respectively, which shows the increasing significance of big data in healthcare. These terms indicate the substantial volumes of health data being generated and analyzed to extract valuable insights for enhancing patient care and public health strategies. Cloud computing with total link strengths. While cloud computing provides scalability and flexibility for managing health data that ensure cloud security in maintaining the confidentiality and integrity of this data. The keywords COVID-19 and COVID-19 pandemic hold total link strengths of 208 and 20 respectively, reflecting the profound impact of the pandemic on healthcare practices. This has led to a rapid adoption of digital health technologies to support remote care and monitoring emphasizing the need for innovative solutions in healthcare informatics. Healthcare is increasingly using machine learning algorithms for tasks such as disease prediction, image analysis, and personalized treatment recommendations which is transforming healthcare informatics. Privacy is another important keyword with a total link strength of 236 this shows the need to protect patient privacy in the digital era. Maintaining privacy in EHR systems is crucial for preserving patient trust and complying with data protection regulations. The security keyword has a total link strength of 246 this shows the importance of protecting EHR systems from unauthorized access and data breaches. Security measures like encryption and access controls are crucial for safeguarding patient information. Trust is a keyword with a total link strength of 48, emphasizing the importance of building trust among patients' healthcare providers and other stakeholders in the healthcare ecosystem. Trust is essential for promoting the adoption and use of digital health technologies. Other keywords such as telehealth, telemedicine, usability, electronic medical record (EMR) are keywords that appear to have a less total link strengths. These findings therefore answer RQ2 and in Fig. 7, most frequently used keywords in EHR are clearly demonstrated.

5 Limitation of Study

With evolving research trends in electronic health record implementation, the search was limited to the Scopus database, and the co-occurrence analysis of relevant terms covered 83% of the total articles. This limitation was primarily due to missing author keyword information from specific journals. To enhance the comprehensiveness of future studies, it is recommended to compare outputs from various databases.

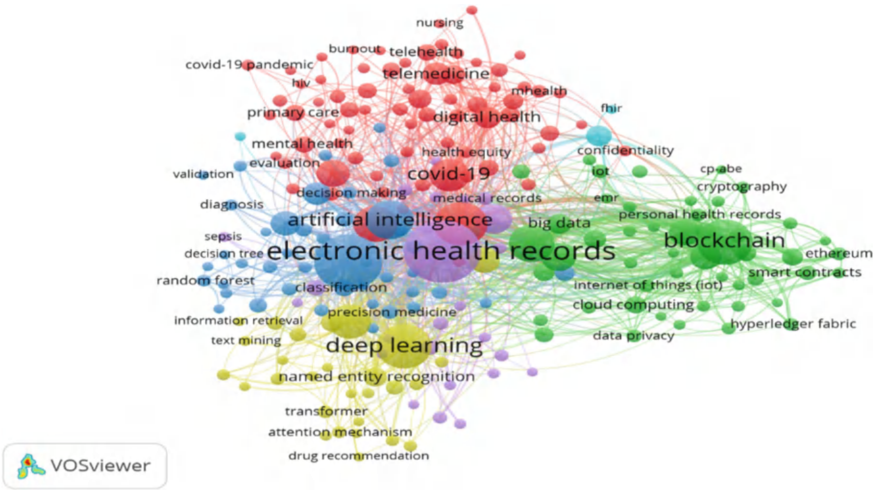


Fig. 7 Co-occurrence network of keywords in electronic health record publications

6 Conclusion

In conclusion, this study conducted a comprehensive bibliometric analysis of EHR research trends spanning from 2014 to 2024. RQ1 highlighted that the field has experienced substantial growth over the past decade, with further expansion expected. The USA emerges as the primary contributor, accounting for over 35% of publications, making it the leading country in this domain. China and India have also significantly contributed, fostering robust international collaborations. This growth opens opportunities for researchers from other nations, including those in Africa and other developing regions, to participate in extensive research collaborations. The study covers a broad spectrum of thoroughly explored areas within EHR.

RQ2 conversely, has identified most frequently used keywords over the decade. EHR obviously take the lead, having the highest link strength. Machine learning and blockchain followed demonstrating the growing integration of state-of-the-art technologies in EHR research. In addition, it also signifies multifaceted nature of EHR research, spanning technological innovation, data management, and ethical considerations.

The findings are important as they give a detailed look at the EHR research paradigm, uncovering the emerging trends and areas of focus. Consequently, it is useful in guiding future research directions, informing policy decisions and indicating possible areas of international collaboration within health informatics which is a fast-growing field. Therefore, this analysis will provide researchers, institutions and policy makers with insight into the current state of EHR research that will help them make better informed decisions about where to put their resources and prioritize on what for future health technology investments.

7 Future Work Direction

Based on this study its recommended that future studies in the field of EHR should focus on these areas as stated:

Track New Trends: It would be great to do a follow-up study to see how the trends in EHR research continue to evolve. By identifying new themes, technologies, and methods that have emerged, we can stay ahead of the curve and understand the direction this field is heading.

Compare Different Regions: It could be really insightful to expand the analysis to compare EHR research trends across different countries or regions. Understanding regional differences, collaborations, and unique challenges can help develop solutions that are tailored to the specific needs of healthcare systems around the world.

Examine Policy Impacts: Looking into how major policy changes and healthcare regulations have influenced EHR research could be valuable. Analyzing the connection between policy shifts and research output, funding patterns, and focus areas can provide a broader understanding of these changes' impacts on the field.

Integrate Emerging Technologies: Exploring how EHR research intersects with other emerging healthcare technologies like AI, telemedicine, and wearable devices could open up exciting new possibilities. This approach can identify areas where these technologies can work together to create more integrated and efficient healthcare solutions.

Focus on User Experience: Diving deeper into how healthcare providers and patients experience EHR systems would be beneficial. Studying usability, satisfaction, and the impact of EHRs on clinical workflows and patient outcomes can provide practical insights for improving these systems to better serve users.

Address Ethical and Privacy Concerns: Investigating how ethical and privacy issues have been handled in EHR research over the past decade is crucial. Future work could focus on managing these concerns and proposing new frameworks to ensure the ethical use of EHR data, safeguarding patient privacy while advancing research.

References

1. Ausserhofer, D., Favez, L., Simon, M., & Zúñiga, F. (2021). Electronic health record use in swiss nursing homes and its association with implicit rationing of nursing care documentation: multicenter cross-sectional survey study. *JMIR Medical Informatics*, *9*(3), 1–14. https://doi.org/10.2196/22974
2. Ayala Solares, J. R., Diletta Raimondi, F. E., Zhu, Y., Rahimian, F., Canoy, D., Tran, J., Pinho Gomes, A. C., Payberah, A. H., Zottoli, M., Nazarzadeh, M., Conrad, N., Rahimi, K., & Salimi-Khorshidi, G. (2020). Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *Journal of Biomedical Informatics*, *101*(March 2019), 103337. https://doi.org/10.1016/j.jbi.2019.103337

3. CHELLADURAI, M. U., Pandian, D. S., & Ramasamy, D. K. (2021). A blockchain based patient centric electronic health record storage and integrity management for e-Health systems. *Health Policy and Technology*, *10*(4), 100513. https://doi.org/10.1016/j.hlpt.2021.100513
4. Doi, S., Citation, U., & Denmark, S. (2018). *University of Southern Denmark The application of bibliometric analysis : disciplinary and user aspects*. https://doi.org/10.1007/S11192-018-2765-Z
5. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Marc, W. (2021). How to conduct a bibliometric analysis : An overview and guidelines. *Journal of Business Research*, *133*(May), 285–296. https://doi.org/10.1016/j.jbusres.2021.04.070
6. Eck, N. J. Van, & Waltman, L. (2023). *VOSviewer Manual*. *October*.
7. Everett-Thomas, R., Joseph, L., & Trujillo, G. (2021). Using virtual simulation and electronic health records to assess student nurses' documentation and critical thinking skills. *Nurse Education Today*, *99*(January), 104770. https://doi.org/10.1016/j.nedt.2021.104770
8. Goldstein, R. L., Anoshiravani, A., Svetaz, M. V., & Carlson, J. L. (2020). Providers' Perspectives on Adolescent Confidentiality and the Electronic Health Record: A State of Transition. *Journal of Adolescent Health*, *66*(3), 296–300. https://doi.org/10.1016/j.jadohealth.2019.09.020
9. Hassett, M. J., Cronin, C., Tsou, T. C., Wedge, J., Bian, J., Dizon, D. S., Hazard-Jenkins, H., Osarogiagbon, R. U., Wong, S., Basch, E., Austin, T., McCleary, N., & Schrag, D. (2022). eSyM: An Electronic Health Record–Integrated Patient-Reported Outcomes–Based Cancer Symptom Management Program Used by Six Diverse Health Systems. *JCO Clinical Cancer Informatics*, *6*, 1–10. https://doi.org/10.1200/cci.21.00137
10. Hathaliya, J. J., & Tanwar, S. (2020). An exhaustive survey on security and privacy issues in Healthcare 4 . 0. *Computer Communications*, *153*(January), 311–335. https://doi.org/10.1016/j.comcom.2020.02.018
11. Ibrahim, M., Kani, Y., & Ahmed, E. (2019). Examining the impact of implementation of electronic health record system for effective health management in Katsina State Hospitals, Nigeria. *International Journal of Gastroenterology*, *3*(2), 27–34. https://doi.org/10.11648/j.ijg.20190302.11
12. Jabali, A. K., Waris, A., Khan, D. I., Ahmed, S., & Hourani, R. J. (2022). Electronic health records: Three decades of bibliometric research productivity analysis and some insights. *Informatics in Medicine Unlocked*, *29*(October 2021), 100872. https://doi.org/10.1016/j.imu.2022.100872
13. Keshta, I., & Odeh, A. (2021). Security and privacy of electronic health records: Concerns and challenges. *Egyptian Informatics Journal*, *22*(2), 177–183. https://doi.org/10.1016/j.eij.2020.07.003
14. Lewkowicz, D., Wohlbrandt, A., & Boettinger, E. (2020). Economic impact of clinical decision support interventions based on electronic health records. *BMC Health Services Research*, *20*(1), 1–13. https://doi.org/10.1186/s12913-020-05688-3
15. Lim, W. M. (2024). *Guidelines for interpreting the results of bibliometric analysis : A sensemaking approach*. 17–26. https://doi.org/10.1002/joe.22229
16. Melnick, E. R., Dyrbye, L. N., Sinsky, C. A., Trockel, M., West, C. P., Nedelec, L., Tutty, M. A., & Shanafelt, T. (2020). The Association Between Perceived Electronic Health Record Usability and Professional Burnout Among US Physicians. *Mayo Clinic Proceedings*, *95*(3), 476–487. https://doi.org/10.1016/j.mayocp.2019.09.024
17. Mishra, V., Liebovitz, D., Quinn, M., Kang, L., Yackel, T., & Hoyt, R. (2022). Factors that influence clinician experience with electronic health records. *Perspectives in Health Information Management*, *19*(1).
18. Nguyen, L., Bellucci, E., & Nguyen, L. T. (2014). Electronic health records implementation : An evaluation of information system impact and contingency factors. *International Journal of Medical Informatics*, *83*(11), 779–796. https://doi.org/10.1016/j.ijmedinf.2014.06.011
19. O'Brien, E. C., Raman, S. R., Ellis, A., Hammill, B. G., Berdan, L. G., Rorick, T., Janmohamed, S., Lampron, Z., Hernandez, A. F., & Curtis, L. H. (2021). The use of electronic health records for recruitment in clinical trials: a mixed methods analysis of the Harmony Outcomes Electronic

Health Record Ancillary Study. *Trials*, *22*(1), 4–11. https://doi.org/10.1186/s13063-021-05397-0

20. Shi, S., He, D., Li, L., Kumar, N., Khan, M. K., & Choo, K. K. R. (2020). Applications of blockchain in ensuring the security and privacy of electronic health record systems: A survey. *Computers and Security*, *97*. https://doi.org/10.1016/j.cose.2020.101966
21. Tanwar, S., Parekh, K., & Evans, R. (2020). Blockchain-based electronic healthcare record system for healthcare 4.0 applications. *Journal of Information Security and Applications*, *50*. https://doi.org/10.1016/j.jisa.2019.102407
22. Tian, H., & Chen, J. (2022). A bibliometric analysis on global eHealth. *Digital Health*, *8*. https://doi.org/10.1177/20552076221091352
23. van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3
24. van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. *Measuring Scholarly Impact*, 285–320. https://doi.org/10.1007/978-3-319-10377-8_13

Performance Analysis of Deep Learning Models in Detection and Counting of Bees for Hive Monitoring

James Alfred Baluran, Dan Lloyd Lomongo, Ivant Fred Ondoy, Jinky Marcelo, Donah Rae Verula, Viceliogems Verula, Mohana Shanmugam, and Lemuel Clark Velasco

Abstract The decline in bee population poses a significant threat to agricultural productivity and biodiversity, highlighting the need for advanced monitoring techniques. While deep learning techniques have shown remarkable potential for object detection tasks, their comparative performance have not been fully examined for the specific application of detecting and counting bees, especially in challenging environments where occlusion and small objects are common. This paper aims to conduct a performance analysis of deep learning-based object detection models for bee detection and counting to monitor hive conditions in Philippine bee farm. The study collected 3050 video clips, resulting in 58,000 frames, from which 5000 images with 11,775 annotations were pre-processed into the BukBee dataset. This dataset serves as the foundation for training and evaluating five advanced object detection models: SSD, Faster R-CNN, YOLOv4, EfficientDet, and CenterNet. The experimental results show that Yolov4 achieved the highest mAP score of 97.9 while SSD, Faster RCNN, EfficientDet, and CenterNet exhibited mAP scores of 82.3, 82.2, 73.3, and 79.7 respectively. The experiments reveal that the best performing model, Yolov4, demonstrated superior capabilities for bee detection and counting, even in challenging scenarios like occlusion and presence of small objects. The performance analysis presented in this study can be integrated into an IoT framework for real-time monitoring of bee hives, thereby promoting sustainable practices in bee farming.

J. A. Baluran · D. L. Lomongo · I. F. Ondoy · J. Marcelo (✉) · D. R. Verula · V. Verula
Information Technology Department, College of Information Sciences and Computing, Central Mindanao University, Maramag, Bukidnon, Philippines
e-mail: jinkymarcelo@cmu.edu.ph

M. Shanmugam
Department of Informatics, College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Putrajaya Campus, Putrajaya, Malaysia

L. C. Velasco
Department of Information Technology, College of Computer Studies, Mindanao State University-Iligan Institute of Technology, Iligan City, Philippines

Premiere Research Institute of Science and Mathematics, Center for Computational Analytics and Modelling, MSU-IIT, Iligan City, Philippines

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_4

Keywords Smart bee farming · Automatic hive monitoring · Bee detection and counting

1 Introduction

Bee farming plays a crucial role in agriculture by supporting crop pollination, global food security and ecosystem biodiversity [1–3]. However, beekeeping faces significant challenges, including colony collapse disorder, pesticide use, parasites, and climate change, all of which contribute to the decline in bee populations [1, 2]. This decline directly threatens honey production and crop pollination, emphasizing the importance of effective hive monitoring to ensure health and sustainability of bee colonies [1, 2, 4, 5].

In the Philippines, traditional beekeeping practices have long relied on manual monitoring techniques to assess hive health and bee populations. This process typically involves physically opening hives to estimate the number of bees and check for signs of disease or other issues [4, 5, 8, 9]. While this method has been used for generations, it is labor-intensive, time-consuming, and prone to human error. The invasive nature of these inspections can also stress the bees, potentially harming their health [6, 10]. Therefore, the limitations of traditional methods highlight the need for more advanced and efficient bee monitoring solutions.

In recent years, technological advancements, particularly in the fields of deep learning and computer vision, have shown great promise in addressing these challenges and improve beekeeping practices [1, 2, 8, 11]. Automated bee detection and counting systems, leveraging techniques like convolutional neural networks (CNNs), can provide real-time data on hive activity, enabling beekeepers to make informed decisions and respond quickly to emerging issues [2]. Additionally, remote monitoring capabilities can allow beekeepers to track their hives from afar, reducing the need for frequent physical inspections [4, 7, 10]. Despite the promise of these technologies, to the best of our knowledge, there are very few studies that have specifically applied deep learning to bee detection and counting, particularly within the Philippine context [3]. Developing models for automatic bee detection and counting is challenging due to occlusion, small size of bees, and lack of localized datasets. Occlusion caused by large numbers of bees presents a significant challenge to achieving accurate detection and counting. Furthermore, the small size and rapid movement of bees pose considerable challenge in detection and counting, necessitating advanced deep learning techniques capable of handling small object detection effectively [1, 7, 17, 18]. The limited availability of localized datasets, particularly in the Philippine context, adds to these challenges, as existing models generally depend on standard architectures like YOLOv4, Faster R-CNN, SSD, EfficientDet, and CenterNet, which may not be ideally suited for the specific requirements of bee detection. Given the variability in performance across different models, it is essential to conduct a comparative evaluation to determine the most accurate and efficient model for real-time bee monitoring, particularly within the context of Philippine bee farm.

This study aims to address the current research gap by conducting a performance analysis of deep learning-based object detection models within the Philippine context. Specifically, the study aims to contribute an image dataset on bees and identify the most effective model for real-time bee monitoring, enhancing the accuracy and efficiency of hive health assessments. By doing so, this study will provide an important tool for beekeepers, enabling them to monitor hive conditions more effectively and make timely, informed decisions to ensure the health and productivity of their colonies.

2 Methodology

The study proposes a deep learning-based approach for automatic bee detection and counting in the Philippines environmental setting. As shown in Fig. 1, the method leverages object detection using deep neural networks [15, 18–21]. Object detection is a computer vision task in which the goal is to detect and locate objects of interest in an image or video [22, 23]. The task involves identifying the position and boundaries of objects in an image, and classifying the objects into different categories. To achieve this, the method comprises data collection, data processing, data annotation, model development, and model evaluation.

2.1 Data Collection

The data collection process for capturing bee activity at the hive entrance was designed using a Raspberry Pi 4 Model B and a Logitech camera, as illustrated in Figs. 2 and 3. The setup involved configuring a camera to record the video at 1280 × 720 resolution with a frame rate of 60 frames per second. This configuration was maintained for 12 h a day to ensure consistent data capture. Figure 2 outlines the setup procedure, positioning the camera 16 inches above the hive entrance at 40° angle. This setup was specifically chosen to reduce occlusion and capture clear footage of bees as they entered and exited the hive Video recording involved capturing 15-s clips every 15 min, which systematically documented bee activity throughout the day.

Following data capture, as shown in Fig. 3, the video files were saved to an SD card within the Raspberry Pi. The SD card was then transferred to a computer,

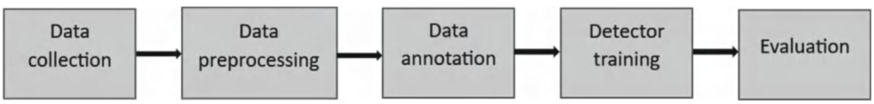


Fig. 1 Overview of the research method

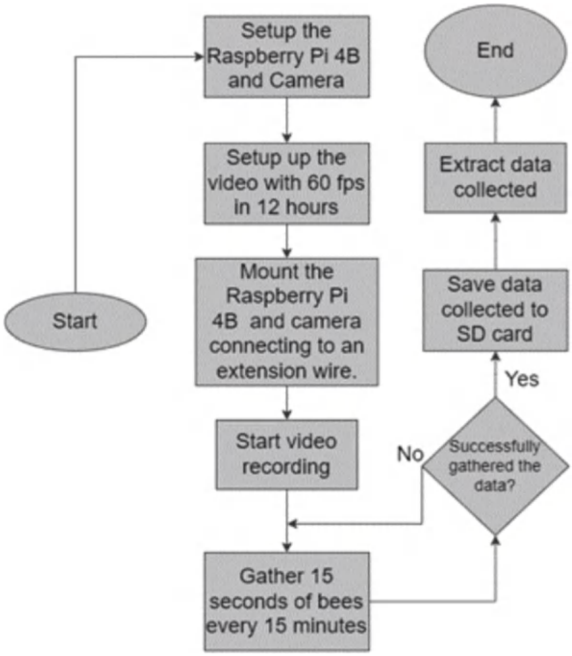


Fig. 2 Flowchart of collecting the bee dataset from setting up the camera to extracting the images

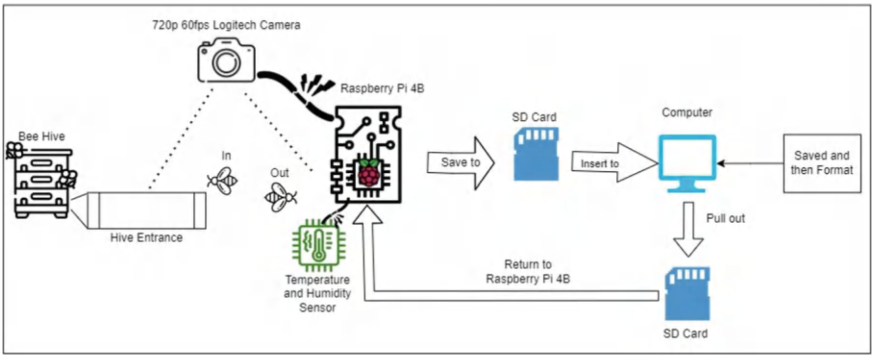


Fig. 3 Data collection method from a bee hive using a 720p camera connected to a Raspberry Pi 4B

where a Python script utilizing OpenCV was executed to automatically extract frames from the video files. These frames were saved for subsequent analysis. After data extraction, the SD card was formatted and reinserted into the Raspberry Pi, ready for the next data collection cycle. This process ensured that high-quality frames were selected for further processing and model training.

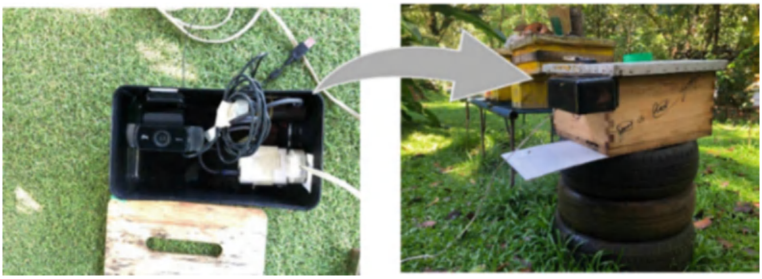


Fig. 4 Dataset collection setup of the study

Fig. 5 Sample images collected in BukBee dataset

Figure 4 illustrates the data collection setup within a Langstroth beehive at the bee farm nursery, showcasing the IoT-based configuration implemented for an experiment conducted at Bukidnon Bee Farm, Bangcud, Malaybalay City, Bukidnon, The Philippines. Notably, the hardware components, including the Raspberry Pi 4 Model B, Logitech c922 Pro web camera, cooling fan, DHT-22 sensor, power supply, and white landing platform, were arranged well for optimal functionality and data collection. The data collection lasted for 5 days continuously, from 5:00 AM to 7:00 PM with a 15 min gap in between 15-s video clips captured in Bukidnon Bee Farm. Data collection had a total of 3,050 15-s video clips. Some of the sample images in BukBee dataset are shown in Fig. 5. The image consists of three side-by-side photos, each showing multiple bees clustered closely together on a flat, light-colored surface, near a hive entrance. The bees in these images are notably small and occluded, with some bees overlapping or partially covering others.

2.2 Data Preprocessing

Preprocessing is an important phase to prepare the captured video data for training the deep learning-based object detection models. Initially, videos of bees entering and leaving the hive that were recorded using the Raspberry Pi 4 Model B and Logitech camera were saved as a series of frames. These frames were then extracted from the video using VLC Media Player. From the collected 3,050 15-s video clips, 375 were randomly selected for extraction of frames. The collection had an original dataset consisting of 58,000 frames with a dimension of 854 $\times$ 480. Subsequently, a random sample of 10,000 frames was obtained. Following the data collection process, the images were subjected to a thorough cleaning, cropping, and gamma correction procedure. Gamma correction was applied to enhance the visibility of darker images, ensuring full detail extraction. Subsequently, the images underwent resizing to a standardized dimension of 640 × 640 and automatic orientation using the preprocessing tool of Roboflow [24–27]. After undergoing cleaning, gamma correction and cropping processes, a total of 5,000 images were pre-processed into the so-called BukBee dataset.

2.3 Data Annotation

Annotation is a vital step in the development of object detection models, as it defines the ground truth that the model will learn from. The accuracy of these annotations directly influences the ability of the models to detect objects in new images [50]. Annotation involves the labeling of regions of interest (ROIs) in images, which in this study, focused on identifying and marking the entrance and landing platform of the beehive where bee activity is concentrated. Using the Roboflow annotation tool [24], bounding boxes were precisely drawn around each bee and the relevant areas of interest. These bounding boxes store the x and y coordinates that define the location of each bee within the image, which are crucial for the object detection model to learn how to identify and localize bees in new, unseen images.

The Roboflow interface, as depicted in Fig. 6, illustrates the process with bees enclosed within yellow bounding boxes, which identify the areas for object detection. The "Labels" section in the interface categorizes each detected object, while the tools panel provides various annotation features, including the Label Assist function. This feature was particularly beneficial for the experiment, as it used a custom model trained specifically for this study to automate and expedite the annotation process [25–27]. After the bounding box annotations were completed, the dataset was partitioned into training and testing sets using an 80:20 split. This resulted in 80% of the images being used for training, while the remaining 20% were set aside for testing.

During the annotation process, it is essential for the study to thoroughly inspect the health and details of the dataset. The dataset has a total of 11,775 annotations. On average, there are 4.9 bounding boxes per image, and the dataset is labeled with a single class, which is "bee". It has been observed that the most frequent count of bees annotated in the dataset is between 2 and 4 per image, as demonstrated in Fig. 7. It is important to note that the bee dataset exhibits instances of occlusion and contains small objects.

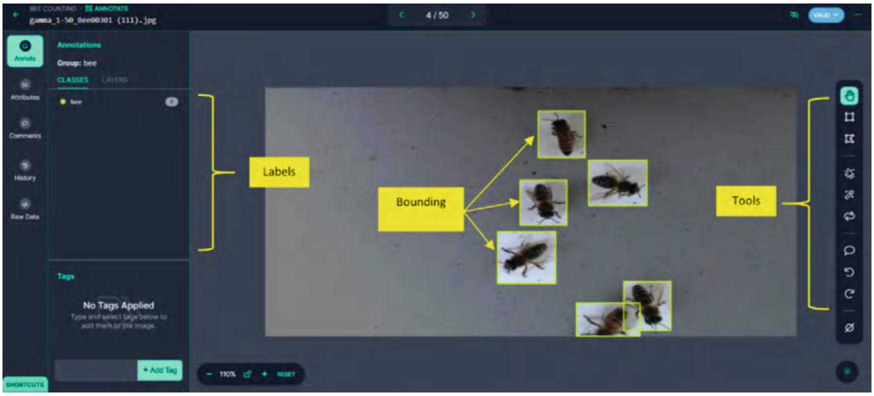


Fig. 6 Roboflow annotation tool interface [29]

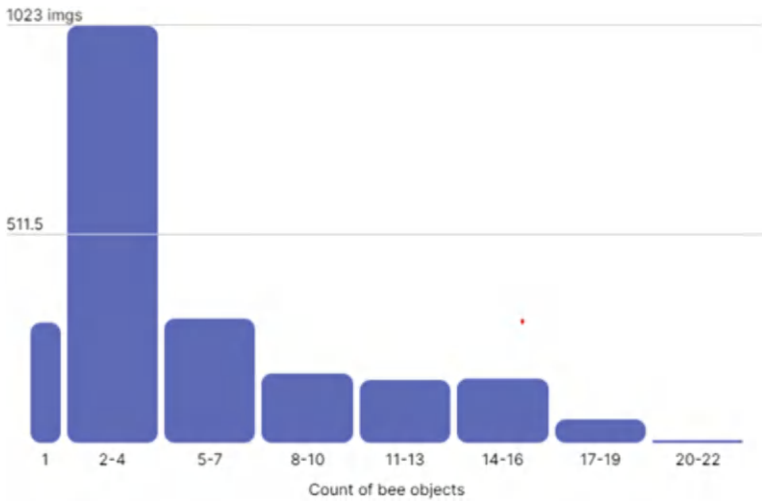


Fig. 7 Histogram of object count by image

2.4 Model Development

During the training and testing phases, the image dataset was divided into 80% for training and 20% for testing. The training dataset is used to teach the model, allowing it to learn patterns, features, and relationships within the data, while the testing dataset is used to assess how well the model generalizes to new, unseen data [56]. To train models for detecting bees, a transfer learning technique was employed [29–32]. This approach involves adapting models that have undergone pre-training on extensive, diverse datasets in similar computer vision tasks to the specific bee counting task. This

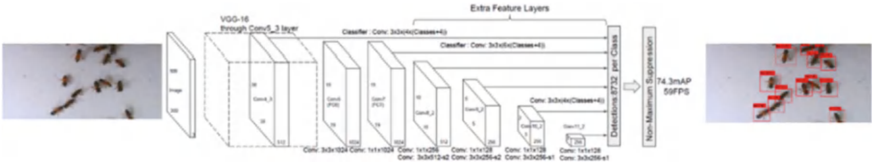


Fig. 8 Single shot multibox network architecture [15]

approach significantly accelerates the training process and improves initial accuracy, allowing the models to effectively identify bees, even in complex environments.

In this study, five state-of-the-art object detection models were utilized, each with unique strengths and capabilities tailored for complex computer vision tasks such as bee detection and counting. The first one was the single shot multibox detector (SSD) it derived its name from the term "single shot" because it detected objects using one forward pass through the CNN as shown in Fig. 8 [15]. Using SSD for detecting and counting bee traffic is beneficial for its fast inference speed and has been utilized in a wide range of applications [15, 33, 34].

Faster R-CNN, also known as Region-Based Convolutional Neural Network, is an improved version of R-CNN, a type of CNN primarily used for processing video data [18]. Unlike SSD, Faster R-CNN operates in two stages. First, it learns to recognize and locate objects within an image frame. It accomplishes this by passing through a fully convolutional network and a region proposal network (RPN), which generates a series of bounding boxes of various sizes and scales across the image [18, 35–37]. These bounding boxes are then inputted into a detection network consisting of fully connected layers, which identifies both labels and offsets, as illustrated in Fig. 9. Using Faster R-CNN for bee detection and counting is advantageous because its two-stage approach and precise object localization help accurately identify and count bees in complex visual environments.

You Only Look Once version 4 (YOLOv4) signifies a significant leap forward in object detection algorithms, known for its remarkable balance between speed and accuracy [19]. This efficiency is particularly crucial in time-sensitive applications like autonomous driving and surveillance. YOLOv4 distinguishes itself with optimizations tailored for GPU utilization, thereby enhancing its computational accessibility. It integrates cutting-edge techniques such as weighted-residual-connections and mish-activation, contributing to its superior performance [19, 38–42]. Moreover, its utilization of CSPDarknet53 as a backbone architecture facilitates efficient feature extraction, rendering YOLOv4 adaptable and scalable across varying object dimensions and dataset sizes, as depicted in Fig. 10. The utilization of YOLOv4 for bee detection and counting provides the advantage of rapid and precise analysis, making it well-suited for real-time monitoring of bee populations.

The EfficientDet model architecture, depicted in Fig. 11, derives from the EfficientNet architecture, emphasizing high accuracy alongside computational efficiency [20]. It integrates additional layers, such as feature pyramid networks (FPN), while optimizing resolution, depth, and width concurrently, thus advancing upon previous

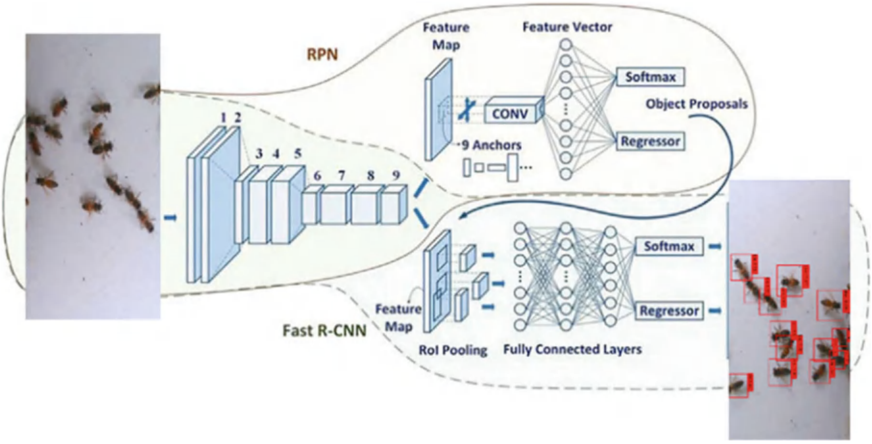


Fig. 9 Faster R-CNN network architecture [18]

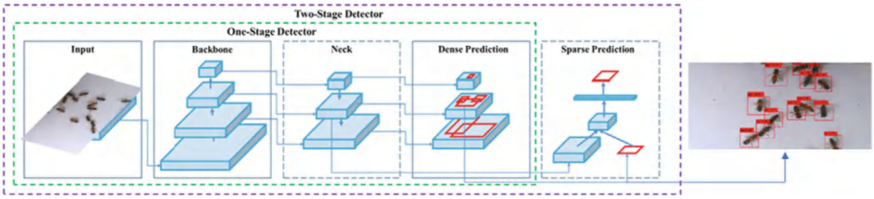


Fig. 10 YOLOv4 network architecture [19]

object detection frameworks like YOLO and Faster R-CNN [20, 43–45]. Pre-trained with the common objects in context (COCO) dataset, this architecture exhibits versatility across a multitude of real-world scenarios requiring object detection. Leveraging the EfficientDet model for bee detection and counting offers the advantage of robust and efficient analysis, well-suited for real-time monitoring and management of bee populations.

CenterNet, a method for anchor-free object detection similar to YOLO, utilizes non-maximum suppression (NMS) to generate key points and heatmaps, which are

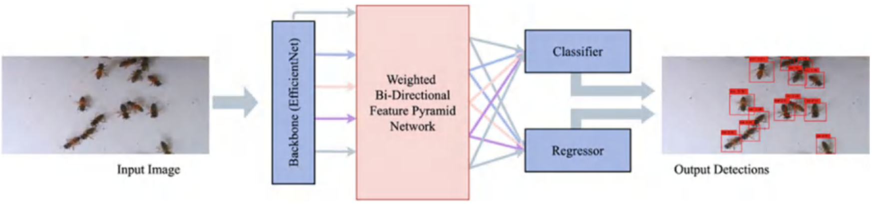


Fig. 11 EfficientDet network architecture [20]

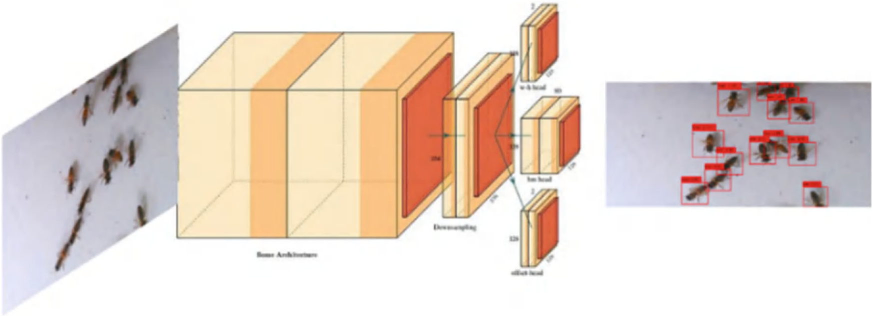


Fig. 12 CenterNet network architecture [21]

then used to derive bounding boxes and predict corresponding object categories, as depicted in Fig. 12 [21]. Notably, CenterNet leverages its key points to estimate object coordinates, scale, and orientation. Despite some shared limitations with YOLO, such as computational cost, CenterNet distinguishes itself as a two-stage detection algorithm. Its resource-efficient design, eliminating the need for anchors and computationally intensive NMS, makes it particularly advantageous for developers monitoring bees with limited computational resources [21, 46–48]. This efficiency renders CenterNet a promising option for bee detection and counting tasks, offering the potential for cost-effective and scalable solutions in bee population monitoring.

The training of the five object detection models—SSD, Faster R-CNN, YOLOv4, EfficientDet, and CenterNet—was conducted using TensorFlow 2 on a high-performance GPU setup. All models were trained on an Intel GeForce RTX 4090 GPU, providing the necessary computational power for efficient model training. Each model was fine-tuned with specific hyperparameters to optimize performance in the task of bee detection and counting. The key hyperparameters used for training each of the five object detection models in this study are summarized in Table 1. These parameters include batch size, optimizer type, learning rate, the number of training steps, and the application of early stopping to prevent overfitting.

Table 1 Summary of the hyperparameters for each network

Model	Batch size	Optimizer	Learning rate	Steps	Early stopping
SSD	8	Momentum	0.08	50,000	Yes
Faster R-CNN	8	Momentum	0.04	25,000	Yes
Yolov4	8	Adam	0.0001	10,000	Yes
EfficientDet	8	Momentum	0.08	300,000	Yes
CenterNet	8	Adam	0.001	140,000	Yes

Table 2 Comparison of BukBee dataset with other bee image datasets

Dataset	Dimension	Quantity
BukBee (ours)	854 × 400	5,000
[8]	3840 × 2160	4,320
[19]	1920 × 1080	2,300
[18]	500 × 200	40

2.5 Evaluation

This study evaluated the trained models for detecting and counting bees using four widely used metrics: precision, recall, F1 score, and mean average precision (mAP) score [49, 50]. Precision measures the proportion of correctly predicted positive instances among all instances identified as positive, providing insight into the model's ability to avoid false positives [49]. Meanwhile, recall quantifies the model's capability to identify all relevant instances, representing the proportion of correctly identified instances among all actual positive instances. F1 score, the harmonic mean of precision and recall, balances the trade-off between precision and recall, offering a comprehensive evaluation of a model's performance, particularly pertinent in scenarios with class imbalance [49]. Lastly, the mean average precision (mAP) score, provides a balanced assessment of the precision-recall curve across varying confidence thresholds, capturing the overall effectiveness of the model [50].

3 Results and Discussion

3.1 BukBee Dataset

Table 2 presents the comparison of BukBee dataset with other existing bee image datasets, providing insights into their dimensions and quantity [11, 16, 51]. The BukBee dataset, developed as part of this study, comprises 5,000 images, making it the largest dataset in terms of quantity. These datasets serve as benchmark datasets for advancing the development of object detection models, such as bees, particularly in various environmental contexts like occlusion and presence of small objects.

3.2 Model Performance Analysis

This section examines the performance of each deep learning technique, including SSD, Faster R-CNN, YOLOv4, EfficientDet and CenterNet using the BukBee dataset. Each model's performance is evaluated based on key metrics: Precision, Recall, F1 Score, mAP Score, and Inference Time. These metrics provide a comprehensive assessment of each model's capability to detect and count bees accurately and efficiently. Tables 3, 4, 5, 6 and 7 present the evaluation results for five deep neural network models utilized in the detection and counting of bees.

Table 3 presents the performance metrics of an SSD model tailored for bee detection and counting. The model achieved a mAP score of 82.3%, which is a critical indicator of its overall detection accuracy since it captures the balance between precision and recall across different confidence thresholds. The precision score of 98.9% further underscores the model's effectiveness in correctly identifying bees, with a low incidence of false positives, while the recall score of 85.9% suggests that while the model is effective, there is still some room for improvement in capturing all bee

Table 3 SSD model for bee detection and counting

Batch size	Optimizer	Learning rate	Steps	Precision	Recall	F1 score	mAP score	Inference time
8	Momentum	Cosine_ decay, base = .08	50000	98.9	85.9	91.8	82.3	0.02 s

Table 4 Faster RCNN model for bee detection and counting

Batch size	Optimizer	Learning rate	Steps	Precision	Recall	F1 score	mAP score	Inference time
8	Momentum	Cosine_ decay, base = 0.04	25,000	98.9	86.4	92.4	82.2	0.24 s

Table 5 Yolov4 model for bee detection and counting

Batch size	Optimizer	Learning rate	Steps	Precision	Recall	F1 score	mAP score	Inference time
8	Adam	Manual_ step, initial = 0.0001	10,000	86	97	91	97.9	0.55 s

Table 6 EfficientDet model for bee detection and counting

Batch size	Optimizer	Learning rate	Steps	Precision	Recall	F1 score	mAP score	Inference time
8	Momentum	Cosine_ decay, base = 8e–2	300,000 (Early stop)	98.7	78.5	87.4	73.3	0.07 s

Table 7 CenterNet model for bee detection and counting

Batch size	Optimizer	Learning rate	Steps	Precision	Recall	F1 score	mAP score	Inference time
8	Adam	Manual_ step, initial = 1e–3	140,000 (Early stop)	98.9	83.3	90	79.7	0.24 s

instances. The F1 score of 91.8% reflects a strong balance between precision and recall, reinforcing the model's capability to perform reliably in practical applications.

The analysis of the Faster R-CNN model training results is presented in Table 4. The Faster RCNN model closely follows with a mAP score of 82.2%, comparable to the SSD model. It maintains a high precision of 98.9% and a slightly better recall of 86.4%, leading to a higher F1 score of 92.4%. This suggests a slightly more balanced detection capability, capturing more bee instances with a similar level of accuracy. However, the inference time of 0.24 s, while slower than SSD, might still be acceptable for many non-real-time applications.

Table 5 shows the performance metrics of a YOLOv4 model tailored for bee detection and counting. The YOLOv4 model exhibits impressive performance in bee detection and counting, highlighted by its high mAP Score of 97.9%. While the precision score of 86% suggests a moderate false positive rate, the recall score of 97% indicates that the model effectively captures a vast majority of bee instances. The balanced F1 score of 91% further validates the model's overall effectiveness, showcasing its ability to harmonize precision and recall in bee detection tasks. Despite a slightly longer inference time of 0.55 s compared to some other models, the YOLOv4 model maintains reasonable efficiency in processing images, ensuring practical feasibility for real-time applications.

Table 6 outlines the performance metrics of an EfficientDet model repurposed for bee detection and counting. The EfficientDet model, with a mAP score of 73.3%, demonstrates lower overall detection accuracy compared to the other models. While it has a high precision of 98.7%, the lower recall of 78.5% suggests that it may miss a significant number of bee instances. This trade-off is reflected in the F1 score of 87.4%. However, its relatively fast inference time of 0.07 s makes it a potential candidate for real-time applications where speed is prioritized over absolute detection accuracy.

Table 7 presents the performance metrics of an CenterNet model finetuned for bee detection and counting. The CenterNet model achieved an mAP score of 79.7%, which is lower than that of the SSD and Faster RCNN models but higher than EfficientDet. With a precision of 98.9% and a recall of 83.3%, it offers a good balance between accuracy and detection completeness, as indicated by the F1 score of 90%. The inference time of 0.24 s suggests that, similar to Faster RCNN, it may be less suited for real-time applications but still provides reliable detection performance.

3.3 Performance Comparison of Models

In this subsection, we present a comparative analysis of the performance of five different object detection models—SSD, Faster R-CNN, YOLOv4, EfficientDet, and CenterNet—trained on the BukBee dataset. Figure 13 shows the visualization of training loss during the training process for five different object detection models: SSD, Faster RCNN, YOLOv4, EfficientDet, and CenterNet. Each subplot displays the plot of training loss over the training iterations or epochs for the respective model on BukBee dataset. The SSD model shows a smooth decay in training loss, indicating stable convergence during training. The Faster RCNN plot exhibits more fluctuations but ultimately converges to a low loss value. EfficientDet shows a more irregular pattern in its training loss curve but still manages to converge. The CenterNet subplot displays a highly oscillatory behavior in the training loss, suggesting potential instability or difficulties in convergence during the training process for this particular model. The YOLOv4 subplot demonstrates a relatively stable training process with the loss gradually decreasing over iterations. Therefore, the SSD, Faster RCNN and YOLOv4 models seem to perform the best in terms of its training stability and convergence behavior while EfficientDet and CenterNet appear to experience more challenges or instabilities during the training phase.

Based on the mAP (mean Average Precision) scores reported for each model, the YOLOv4 model demonstrates the highest mAP score at 97.9%, indicating superior precision and recall balance in detecting and counting bees. This model is closely followed by the SSD and Faster R-CNN models, which both exhibit mAP scores in the low 80 s (82.3% for SSD and 82.2% for Faster R-CNN). Although the SSD model shows slightly better mAP than Faster R-CNN, the difference is marginal, suggesting similar performance levels between these two models in terms of overall detection accuracy. The EfficientDet and CenterNet models, with mAP scores of 73.3% and 79.7% respectively, show lower performance, particularly EfficientDet, which appears to struggle more with achieving a high balance between precision and recall. These results suggest that YOLOv4 is the most effective model for bee detection and counting in terms of mAP, followed by SSD and Faster R-CNN, while EfficientDet and CenterNet demonstrate more variability in performance, likely due to the instability observed in their training processes.

The performance of the models on the BukBee dataset not only surpasses previous benchmarks but also demonstrates significant advancements in detection accuracy and efficiency. For instance, the SSD model achieved a mAP score of 82.3%,

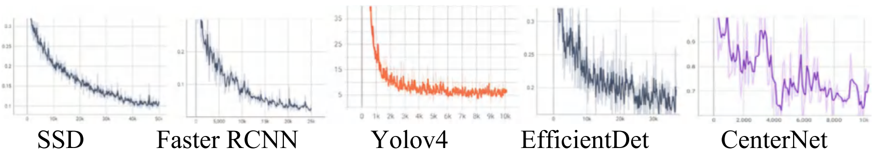


Fig. 13 Comparison of visualization of training loss of the five detection models

markedly higher than the 42.23% and 40.83% reported for MobileNet v1 and Inception v2 SSD models in [12], indicating substantial improvements. The YOLOv4 model excelled even further, with a 97.9% accuracy on a larger dataset of 5,000 images, outperforming similar YOLO-based methods that achieved 87 and 93% accuracy on smaller datasets [2].

3.4 Selection and Deployment of the Best Model

Among the models evaluated for bee detection and counting, YOLOv4 achieves the highest mAP score of 97.9%, indicating superior accuracy in detecting and localizing bee instances within images. mAP is a critical metric in assessing object detection models as it provides a comprehensive evaluation of the performance of the model across different confidence thresholds. It considers both precision and recall, reflecting the ability of the model to accurately detect and localize objects of interest. However, it has a relatively higher inference time of 0.55 s compared to other models. While inference time is a crucial factor, especially in real-time applications, prioritizing mAP ensures reliable monitoring and decision-making regarding bee populations. Despite the relatively higher inference time of certain YOLOv4 models compared to others, their high mAP scores outweigh this drawback in the context of bee detection and counting.

The architecture of YOLOv4 plays a role in achieving its high mAP score in detecting bees, even amidst challenges like occlusion and small objects [19, 39]. The single-shot detection approach of YOLO allows it to process the entire image in one pass through the neural network, capturing global context and spatial relationships that aid in robust detection [19, 38, 42]. Moreover, its grid-based prediction mechanism divides the image into cells, enabling precise localization of bee instances within each cell, even when partially occluded or densely populated. The utilization of anchor boxes further enhances the adaptability of YOLO to diverse sizes and shapes of bee instances, ensuring accurate detection of small objects like bees. Additionally, the integration of a FPN extracts multi-scale features, enabling it to capture fine-grained details of bee instances at different resolutions, enhancing sensitivity to small and occluded objects. Lastly, specialized loss function of YOLO optimizes for both precise localization and accurate classification of detected objects, further contributing to its high mAP score [42].

The result of this study corroborates with the recent literature which highlights the effectiveness of YOLO in bee-related applications, demonstrating its potential for accurate detection and counting tasks in beekeeping and hive monitoring. For example, researchers developed a prototype using YOLOv3 technology to detect the activeness of stingless beehives, comparing SqueezeNet and DarkNet-53 architectures for improved performance [52]. Additionally, another group of researchers employed the YOLO model for detecting and tracking honey bees, showcasing good tracking accuracy and low latency for real-time monitoring systems [53].

Fig. 14 Sample detection of the fine-tuned YOLOv4 in the BukBee testing dataset

The study selected the YOLOv4 model as the most effective for the project due to its superior performance in accurately detecting and counting bee traffic around the hive. As shown in Fig. 14, the fine-tuned YOLOv4 clearly demonstrates its performance in accurately detecting the bees. The images show the model consistently applying precise labels around the bees, evidence of the robust and efficient algorithm at work. This performance is impressive even under different environmental conditions such as occlusion and presence of small objects, aligning perfectly with the model's design which is aimed at handling challenging detection tasks. This is also supported by research, further validating that YOLOv4 is well-suited for practical ecological monitoring and research efforts [54].

4 Conclusion and Recommendations

This study investigated the performance analysis of deep learning models for bee detection and counting to address the need for less invasive and vision-based monitoring of bee hives. The study collected 3050 video clips, resulting in 58,000 frames, from which 5000 images were pre-processed into the BukBee dataset. The BukBee dataset contains 11,775 annotations with an average of 4.9 bounding boxes per image, labeled with a single class 'bee' and exhibiting instances of occlusion and small objects.

The study conducted a thorough evaluation of five advanced object detection models, including SSD, Faster R-CNN, YOLOv4, EfficientDet, and CenterNet, on the 'BukBee' dataset. The result shows that YOLOv4 demonstrated exceptional performance with a high mean average precision (mAP) score of 97.9, effectively detecting bees even in challenging scenarios, while SSD, Faster RCNN, EfficientDet, and CenterNet exhibited mAP scores of 82.3, 82.2, 73.3, and 79.7 respectively, showing varying degrees of effectiveness and highlighting areas for improvement in detection accuracy and efficiency. The study demonstrates that Yolov4 has the superior performance to accurately detect and localize bees, even in challenging scenarios such as occlusion and small object sizes. The findings show the potential of deep learning models towards designing an intelligent monitoring system for bee hives.

While this study has achieved significant progress in automated bee detection and counting, several areas warrant further exploration and research. Expanding the BukBee dataset to include a wider range of environmental conditions, bee species, and hive configurations would enhance the generalizability and robustness of the developed models. Further, incorporating the developed models into an Internet of Things (IoT) framework and deploying them on edge computing devices could enable real-time hive monitoring and facilitate remote data collection and analysis.

References

1. H. C. Nguyen *et al.*, "A method for automatic honey bees detection and counting from images with high density of bees," *2022 IEEE Ninth International Conference on Communications and Electronics (ICCE)*, Nha Trang, Vietnam, 2022, pp. 406–411, https://doi.org/10.1109/ICCE55644.2022.9852024.
2. S. Bilik *et al.*, "Machine learning and computer vision techniques in continuous beehive monitoring applications: A survey," *Comput Electron Agric*, vol. 217, p. 108560, 2024, https://doi.org/10.1016/j.compag.2023.108560.
3. C. R. Cervancia, "Management and conservation of Philippine bees," in *Asian Beekeeping in the 21st Century*, G. and N. P. Chantawannakul Panuwan and Williams, Ed., Singapore: Springer Singapore, 2018, pp. 307–321. https://doi.org/10.1007/978-981-10-8222-1_14.
4. J. Sakanovic Semir and Kevric, "HaBEEtat: A Novel Monitoring Platform for More Efficient Honey Production," in *CMBEBIH 2019*, R. and G. P. L. Badnjevic Almir and Škrbić, Ed., Cham: Springer International Publishing, 2020, pp. 193–200.
5. P. Hristov, R. Shumkova, N. Palova, and B. Neov, "Factors associated with honey bee colony losses: A mini-review," *Vet Sci*, vol. 7, no. 4, 2020, https://doi.org/10.3390/vetsci7040166.
6. M. A. P. Lima, G. C. Cutler, G. Mazzeo, and M. Hrncir, "Editorial: The decline of wild bees: Causes and consequences," *Front Ecol Evol*, vol. 10, 2022, https://doi.org/10.3389/fevo.2022.1027169.
7. R. Tashakkori, A. S. Hamza, and M. B. Crawford, "Beemon: An IoT-based beehive monitoring system," *Comput Electron Agric*, vol. 190, p. 106427, 2021, https://doi.org/10.1016/j.compag.2021.106427.
8. G. Voudiotis, A. Moraiti, and S. Kontogiannis, "Deep Learning Beehive Monitoring System for Early Detection of the Varroa Mite," *Signals*, vol. 3, no. 3, pp. 506–523, 2022, https://doi.org/10.3390/signals3030030.
9. O. Mudrakova, A. Kolodochkin, and Y. Beresneva, "Information system using computer vision technology for innovative beekeeping development," in *BIO Web of Conferences*, EDP Sciences, Jan. 2024. https://doi.org/10.1051/bioconf/20248303006.
10. B. Magnier., G. Ekszterowicz., J. Laurent., M. Rival., and F. Pfister., "Bee hive traffic monitoring by tracking bee flight paths," in *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, SciTePress, 2018, pp. 563–571. https://doi.org/10.5220/0006628205630571.

11. M. Liu *et al.*, "Detection of Varroa destructor Infestation of Honeybees Based on Segmentation and Object Detection Convolutional Neural Networks," *AgriEngineering*, vol. 5, no. 4, pp. 1644–1662, 2023, https://doi.org/10.3390/agriengineering5040102.
12. Voudiotis, G.; Kontogiannis, S.; Pikridas , C. Proposed Smart Monitoring System for the Detection of Bee Swarming. Inventions 2021, 6, 87. https://doi.org/10.3390/inventions6040087
13. J.-S. Ryu, J.-W. Jung, C.-H. Jeong, B.-J. Choi, M. Lee, and H. W. Kwon, "Honeybee In-Out Monitoring System by Object Recognition and Tracking from Real-Time Webcams," *Journal of Apiculture*, vol. 36, no. 4, pp. 273–280, Nov. 2021, https://doi.org/10.17519/apiculture.2021.11.36.4.273.
14. S. Kachole, G. Hunter, and O. Duran, "A computer vision approach to monitoring the activity and well-being of honeybees," 2020, https://doi.org/10.3233/AISE200036.
15. D. and E. D. and S. C. and R. S. and F. C.-Y. and B. A. C. Liu Wei and Anguelov, "SSD: Single shot multibox detector," in *Computer Vision – ECCV 2016*, J. and S. N. and W. M. Leibe Bastian and Matas, Ed., Cham: Springer International Publishing, 2016, pp. 21–37.
16. P. Kongsilp, U. Taetragool, and O. Duangphakdee, "Individual honey bee tracking in a beehive environment using deep learning and Kalman filter," *Sci Rep*, vol. 14, no. 1, Dec. 2024, https://doi.org/10.1038/s41598-023-44718-y.
17. S. M. Williams, N. Aldabashi, P. Cross, and C. Palego, "Challenges in developing a real-time bee-counting radar," *Sensors*, vol. 23, no. 11, 2023, https://doi.org/10.3390/s23115250.
18. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds., Curran Associates, Inc., 2015.
19. A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection", arXiv:2004.10934, Apr. 2020, https://doi.org/10.48550/arXiv.2004.10934
20. M. Tan, R. Pang, and Q. V Le, "EfficientDet: Scalable and Efficient Object Detection," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 10778–10787. https://doi.org/10.1109/CVPR42600.2020.01079.
21. K. Duan, S. Bai, L. Xie, H. Qi, Q. Huang, and Q. Tian, "CenterNet: Keypoint Triplets for Object Detection," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 6568–6577. https://doi.org/10.1109/ICCV.2019.00667.
22. M. Tzampazaki, C. Zografos, E. Vrochidou, and G. A. Papakostas, "Machine Vision—Moving from Industry 4.0 to Industry 5.0," *Applied Sciences*, vol. 14, no. 4, 2024, https://doi.org/10.3390/app14041471.
23. P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, 2001, pp. I–I. https://doi.org/10.1109/CVPR.2001.990517.
24. Dwyer, B., Nelson, J., Hansen, T., et. al. (2024). Roboflow (Version 1.0) [Software]. Available from https://roboflow.com. Computer vision.
25. G. B. Hantos, G. Simon, M. Hejda, A. L. Bernassau, and M. P. Y. Desmulliez, "Automated Particle and Cell Phenotyping Using Object Recognition and Tracking Based on Machine Learning Algorithms," in *2021 IEEE International Ultrasonics Symposium (IUS)*, 2021, pp. 1–4. https://doi.org/10.1109/IUS52206.2021.9593615.
26. E. Elhariri, N. El-Bendary, and S. Mahmoud Saleh, "Strawberry-DS: Dataset of annotated strawberry fruits images with various developmental stages," *Data Brief*, vol. 48, p. 109165, 2023, https://doi.org/10.1016/j.dib.2023.109165.
27. M. Sarnovský and M. Kolárik, "Use of deep learning to automate the annotation of USG lung image data," in *2024 IEEE 22nd World Symposium on Applied Machine Intelligence and Informatics (SAMI)*, 2024, pp. 339–342. https://doi.org/10.1109/SAMI60510.2024.10432851.
28. T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning data mining, inference, and prediction." *Springer Series in Statistics*, 2009.
29. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587. https://doi.org/10.1109/CVPR.2014.81.

30. T. Chen, M. Wu, and H. Li, "A general approach for improving deep learning-based medical relation extraction using a pre-trained model and fine-tuning," *Database (Oxford)*, vol. 2019, Jan. 2019, https://doi.org/10.1093/database/baz116.
31. D. Singh *et al.*, "Classification and analysis of pistachio species with pre-trained deep learning models," *Electronics (Basel)*, vol. 11, no. 7, 2022, https://doi.org/10.3390/electronics11070981.
32. X. Han *et al.*, "Pre-trained models: Past, present and future," *AI Open*, vol. 2, pp. 225–250, 2021, https://doi.org/10.1016/j.aiopen.2021.08.002.
33. A. Arora, A. Grover, R. Chugh, and S. S. Reka, "Real Time Multi Object Detection for Blind Using Single Shot Multibox Detector," *Wirel Pers Commun*, vol. 107, no. 1, pp. 651–661, 2019, https://doi.org/10.1007/s11277-019-06294-1.
34. A. Kumar and S. Srivastava, "Object Detection System Based on Convolution Neural Networks Using Single Shot Multi-Box Detector," *Procedia Comput Sci*, vol. 171, pp. 2610–2617, 2020, https://doi.org/10.1016/j.procs.2020.04.283.
35. B. Liu, W. Zhao, and Q. Sun, "Study of object detection based on Faster R-CNN," in *2017 Chinese Automation Congress (CAC)*, 2017, pp. 6233–6236. https://doi.org/10.1109/CAC.2017.8243900.
36. Q. Fan, L. Brown, and J. Smith, "A closer look at Faster R-CNN for vehicle detection," in *2016 IEEE Intelligent Vehicles Symposium (IV)*, 2016, pp. 124–129. https://doi.org/10.1109/IVS.2016.7535375.
37. H. Jiang and E. Learned-Miller, "Face detection with the faster R-CNN," in *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 2017, pp. 650–657. https://doi.org/10.1109/FG.2017.82.
38. C. Dewi, R.-C. Chen, Y.-T. Liu, X. Jiang, and K. D. Hartomo, "Yolo V4 for Advanced Traffic Sign Recognition With Synthetic Training Data Generated by Various GAN," *IEEE Access*, vol. 9, pp. 97228–97242, 2021, https://doi.org/10.1109/ACCESS.2021.3094201.
39. R. Gai, N. Chen, and H. Yuan, "A detection algorithm for cherry fruits based on the improved YOLO-v4 model," *Neural Comput Appl*, vol. 35, no. 19, pp. 13895–13906, 2023, https://doi.org/10.1007/s00521-021-06029-z.
40. D. Wu, S. Lv, M. Jiang, and H. Song, "Using channel pruning-based YOLO v4 deep learning algorithm for the real-time and accurate detection of apple flowers in natural environments," *Comput Electron Agric*, vol. 178, p. 105742, 2020, https://doi.org/10.1016/j.compag.2020.105742.
41. J. Yu and W. Zhang, "Face mask wearing detection algorithm based on improved YOLO-v4," *Sensors*, vol. 21, no. 9, 2021, https://doi.org/10.3390/s21093263.
42. J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Mach Learn Knowl Extr*, vol. 5, no. 4, pp. 1680–1716, 2023, https://doi.org/10.3390/make5040083.
43. H. Alhichri, A. S. Alswayed, Y. Bazi, N. Ammour, and N. A. Alajlan, "Classification of Remote Sensing Images Using EfficientNet-B3 CNN Model With Attention," *IEEE Access*, vol. 9, pp. 14078–14094, 2021, https://doi.org/10.1109/ACCESS.2021.3051085.
44. J. Wang, L. Yang, Z. Huo, W. He, and J. Luo, "Multi-Label Classification of Fundus Images With EfficientNet," *IEEE Access*, vol. 8, pp. 212499–212508, 2020, https://doi.org/10.1109/ACCESS.2020.3040275.
45. G. Marques, D. Agarwal, and I. de la Torre Díez, "Automated medical diagnosis of COVID-19 through EfficientNet convolutional neural network," *Appl Soft Comput*, vol. 96, p. 106691, 2020, https://doi.org/10.1016/j.asoc.2020.106691.
46. T. Nazir *et al.*, "Detection of diabetic eye disease from retinal images using a deep learning based CenterNet model," *Sensors*, vol. 21, no. 16, 2021, https://doi.org/10.3390/s21165283.
47. R. Wang and C. F. Cheung, "CenterNet-based defect detection for additive manufacturing," *Expert Syst Appl*, vol. 188, p. 116000, 2022, https://doi.org/10.1016/j.eswa.2021.116000.
48. E. and V. D. Xu Zhujun and Hrustic, "CenterNet Heatmap Propagation for real-time video object detection," in *Computer Vision—ECCV 2020*, H. and B. T. and F. J.-M. Vedaldi Andrea and Bischof, Ed., Cham: Springer International Publishing, 2020, pp. 220–234.

49. D. M. W. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *Journal of Machine Learning Technologies*, vol. 2, no. 1, pp. 37–63, 2011, [Online]. Available: https://api.semanticscholar.org/CorpusID:3770261
50. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," *Int J Comput Vis*, vol. 88, no. 2, pp. 303–338, 2010, https://doi.org/10.1007/s11263-009-0275-4.
51. T. Sledevič and D. Matuzevičius, "Labeled dataset for bee detection and direction estimation on entrance to beehive," *Data Brief*, vol. 52, p. 110060, 2024, https://doi.org/10.1016/j.dib.2024.110060.
52. S. I. Safie, N. S. A. Kamal, E. M. M. Yusof, M. Z.-W. M. Tohid, and N. H. Jaafar, "Comparison of squeezeNet and DarkNet-53 based YOLO-V3 Performance for beehive intelligent monitoring system," in *2023 IEEE 13th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, 2023, pp. 62–65. https://doi.org/10.1109/ISCAIE57739.2023.10165285.
53. H. K. Benahmed, M. L. Bensaad, and N. Chaib, "Detection and tracking of honeybees using YOLO and StrongSORT," in *2022 2nd International Conference on Electronic and Electrical Engineering and Intelligent System (ICE3IS)*, 2022, pp. 18–23. https://doi.org/10.1109/ICE3IS56585.2022.10010142.
54. P. Gao *et al.*, "Dynamic Beehive Detection and Tracking System Based on YOLO V5 and Unmanned Aerial Vehicle," *Journal of Biosystems Engineering*, vol. 47, no. 4, pp. 510–520, 2022, https://doi.org/10.1007/s42853-022-00166-6.
55. Everingham, M., Eslami, S. M. A., Van Gool, L., Williams, C. K. I., Winn, J., & Zisserman, A. (2015). The Pascal Visual Object Classes Challenge: A Retrospective. International Journal of Computer Vision, 111(1), 98-136. https://doi.org/10.1007/s11263-014-0733-5
56. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press

Smarter Learning Environment: Exploring the Need for an Intelligent Tutoring System in Iraqi Universities

Ahmed Rashid Alkhuwaylidee, Ali khalaf Gatea, and Mohammed F. Alomari

Abstract The research aims to address the importance of the smart teaching system to guide and assist university students. By improving their learning experience. This system helps solve several problems related to the limited resources of universities and students, language barriers, and cultural differences. Uses technology to provide individual feedback, adaptive learning methods, and data analytics to improve student performance. Key features of the Smart Tutoring system include an easy-to-use interface, adaptive learning methods, and data analytics to monitor student achievement, identify areas for improvement, and help deliver personalized interventions. The system increases student engagement and motivation, improving the learning experience and promoting active participation in education.

Keywords E-learning · Environment · Design architecture · Intelligent tutoring system · Iraq student university

1 Introduction

Iraqi universities face several challenges when it comes to providing quality education to their students. Limited resources, overcrowded classrooms, and a lack of individualized attention often hinder students' learning experiences. This is an area where study systems can play a role in reducing empty space [1–3].

The design architecture of an intelligent tutoring system plays A major role in the growth of a Educational intelligence environment for Iraqi university students. In

A. R. Alkhuwaylidee
University of Thi-Qar, Nasiriyah, Iraq
e-mail: ahmed.alkhuwaylidee@utq.edu.iq

A. Gatea (✉)
Southern Technical University, Basra, Iraq
e-mail: ali.khalaf83@stu.edu.iq

M. F. Alomari
Department of Computer Engineering Techniques, Mazaya University College, Nasiriyah, Iraq

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_5

this study, we explore the key features and design considerations needed to develop an effective system that addresses the specific challenges faced by these students [3]. Firstly, it is important to understand the need for intelligent tutoring systems in Iraqi universities. These systems help students overcome the obstacles facing university education obstacles and enhancing their learning experience [4]. By leveraging technology, these systems can offer tailored feedback, adaptive learning techniques, and data analytics to improve student outcomes. Next, we delve into the challenges faced by Iraqi university students. Limited access to resources, language barriers, and cultural differences can hinder their academic progress [5].

An intelligent tutoring system needs to address these challenges by providing content in multiple languages, incorporating culturally relevant examples, and offering resources that are easily accessible to all students [6].

The key features of an intelligent tutoring system include a user-friendly interface, adaptive learning techniques, data analytics for tracking progress, and gamification for enhancing student engagement and motivation. The design architecture should incorporate data analytics to track student progress, identify areas for improvement, and provide targeted interventions [7–10]. By incorporating game-like elements like badges, leader boards, and rewards, providing effective education requires creating systems that encourage students' skills. Approach ensures a personalized and effective learning journey for students [11].

The research aims to explore the need for an Intelligent Tutoring System (ITS) in Iraqi universities. This system has the potential to revolutionize education by providing individualized and adaptive learning experiences to students [12]. With the rise of e-learning and the increasing demand for accessible and quality education, it is imperative to address the challenges faced by Iraqi university students [13]. This article will delve into the design architecture of an ITS [10], highlighting key features, user interface design, adaptive learning techniques, data analytics, and the integration of gamification [2]. Through this exploration, we aim to paint a comprehensive picture of how an ITS can enhance the educational landscape in Iraq [14].

2 Understanding the Need for Intelligent Tutoring Systems in Iraqi Universities

Iraqi universities have various obstacles in offering quality education to its students. Limited resources, overcrowded classrooms, and a lack of individualized attention frequently impede students' learning experiences. Intelligent tutoring systems can play an important role in bridging the gap [1].

Shortcomings of Traditional Classroom Settings: In a traditional classroom setting, it is difficult for teachers to cater to the diverse learning needs of each student. Some students may grasp concepts quickly, while others may struggle to keep up. Smart systems work to provide guidance that helps in personalized learning experiences according to the students' capabilities [15].

Enhancing Student Engagement: Traditional lectures sometimes fail to engage students effectively. The use of intelligent tutoring systems allows for interactive and dynamic learning experiences that can captivate students' attention and improve their overall engagement in the learning process. For example, gamification techniques can be incorporated to make learning more enjoyable and motivating [2].

Providing Timely Feedback: One of the key advantages of intelligent tutoring systems is the ability to provide immediate and personalized feedback to students. These approaches usually provide further improvements in addressing students' needs and improving their choices over time. Students no longer have to wait for assignment grades or exams to receive feedback; they can get it in real-time [16].

Individualized Learning Paths: Every student has unique strengths, weaknesses, and learning preferences. Intelligent tutoring systems can adapt to each student's needs and create individualized learning paths. For example, if a student is struggling with a particular concept, the system can provide additional practice exercises or resources to reinforce their understanding [17].

Addressing the Accessibility Gap: In some cases, geography or other circumstances may restrict students from accessing quality education. Intelligent tutoring systems Providing the opportunity for distance learning allows for rapid education and development of students' skills in a way that suits them and in a place and time appropriate to their circumstances. This helps bridge the accessibility gap and ensures that all students have equal opportunities to learn [18].

In summary, the need for intelligent tutoring systems in Iraqi universities arises from the challenges faced by traditional classroom settings, the potential to enhance student engagement, the provision of timely feedback, the ability to create individualized learning paths, and the opportunity to bridge the accessibility gap. By addressing these needs, Educational systems based on intelligence can create a revolution in the ways of delivering information to students in a more intelligent and knowledgeable way. Subsequent paragraphs, however, are indented (Fig. 1).

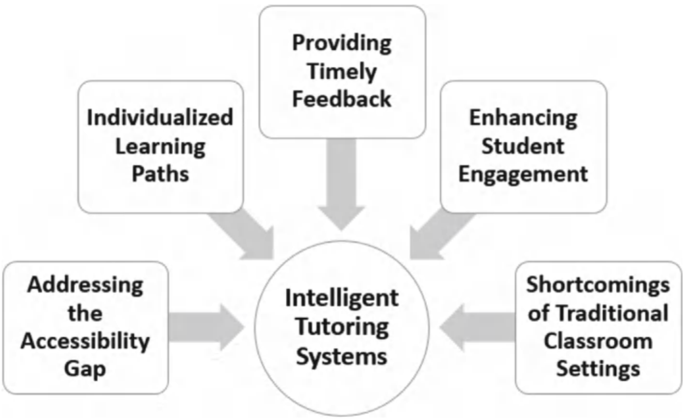


Fig. 1 Understanding the Need for (ITS) in Iraqi Universities

2.1 Exploring the Challenges Faced by Iraqi University Students

Iraqi university students face a multitude of challenges that can hinder their learning experiences. These challenges can arise from various factors such as limited resources, language barriers, and cultural differences. By understanding these challenges, designers of an Smart education systems can tailor the design architecture to address these specific needs and create a more effective learning environment [19].

Limited Resources

One of the major challenges faced by Iraqi university students is the lack of resources, including textbooks, educational materials, and access to technology. This limitation can prevent students from acquiring the necessary knowledge and skills to succeed in their studies. To overcome this challenge, the design architecture of an intelligent tutoring system can provide virtual libraries, online resources, and interactive multimedia content. By incorporating these features, students can have access to a wide range of materials that can supplement their learning [3].

Language Barriers.

Another significant challenge for Iraqi university students is language barriers. English is often the primary medium of instruction in higher education, but many students may not have a strong command of the language. This can lead to difficulties in understanding lectures, reading textbooks, and communicating with instructors. To address this challenge, the design architecture of the intelligent tutoring system can incorporate language support features such as translation tools, vocabulary assistance, and interactive language exercises. These features can help students overcome language barriers and enhance their comprehension and communication skills [20]

Cultural Differences.

Cultural differences can also pose challenges for Iraqi university students. The educational system in Iraq may have different teaching methods, assessment styles, and expectations compared to other countries. This can create a disconnect and make it difficult In order for students to adapt to the atmosphere surrounding teaching. The design architecture of the intelligent tutoring system can mitigate this challenge by incorporating adaptive learning techniques. These techniques can personalize the learning experience based on individual student preferences, learning styles, and cultural backgrounds. By adapting the content and approach to align with the students' cultural context, the system can enhance their engagement and understanding [3].

By exploring and understanding the challenges faced by Iraqi university students, the design architecture of an intelligent tutoring system can be tailored to address these specific needs. By providing resources, language support, and personalized learning experiences, the system can create a smarter learning environment that promotes student success.

2.2 Key Features of an Intelligent Tutoring System

An intelligent tutoring system (ITS) incorporates various features that make it a valuable tool for supporting learning in Iraqi universities. These key features are designed In order to enhance the student's personal experience and guide him towards the best. Here are some essential features of an ITS:

1. Individualized Instruction: One of the primary benefits of an ITS is its ability to provide individualized instruction to students. Through the use of adaptive learning algorithms, the system can analyses a student's performance and tailor the learning materials and activities to their specific needs. By analogy to the above, ITS can provide arguments that explain the student's struggle with a particular concept or problem to improve his knowledge and understanding skills [21].
2. Real-time Feedback: Another crucial feature of an ITS It reflects the possibility of recording direct notes to save time in recording notes in the lesson and reviewing them later and assessments in real-time. This instant feedback allows students to identify their mistakes and areas for improvement, enabling them to make corrections and progress more efficiently [22].
3. Tracking progress: A tracking feature is usually provided for student data in the systems, analyzing strengths, weaknesses, and appropriate patterns of learning. Data can create detailed progress reports, allowing parties involved in the educational process to properly and comprehensively track the stages of development of the student's educational level [23].
4. Adaptive Assessments: Traditional assessments often follow a one-size-fits-all approach, which may not accurately measure a student's true understanding of the material. In contrast, an ITS can provide more effective results in evaluating the skills and knowledge possessed by students. By using more accurate algorithms that match the answers provided in the tests in an interactive manner with the difficulties faced by students [24].
5. Collaborative Learning: An ITS can also facilitate collaborative learning among students. Through features such as discussion boards or virtual group projects, students can engage in peer-to-peer learning and share their knowledge and perspectives. This collaborative environment fosters active participation and encourages students to learn from each other [25] (Fig. 2).

Fig. 2 Key features of an intelligent tutoring system

By incorporating these key features, an intelligent tutoring system can revolutionize the learning experience for Iraqi university students. It provides personalized instruction, immediate feedback, progress tracking, adaptive assessments, and opportunities for Joint learning. With the support of an ITS, students can enhance their understanding, improve their performance, and develop essential skills that will benefit them throughout their academic journey.

2.3 Designing an Effective User Interface for the Intelligent Tutoring System

When it comes to designing an effective user interface for an intelligent tutoring system, simplicity and ease of use should be the top priorities. The interface should be intuitive, allowing students to navigate through the system without any confusion or difficulty. A cluttered and complicated interface can deter students from using the system and hinder their learning experience [26].

One important aspect to consider is the use of clear and concise language in the interface. The instructions and prompts should be written in a way that is easily understandable to the students. For example, instead of using technical jargon or complex terminology, the interface should use simple and everyday language that the students are familiar with. This ensures that students can easily comprehend the information provided by the system and complete their learning tasks without any confusion [27].

Another key element of an effective user interface is visual design. The layout, colors, and typography should be visually appealing and consistent throughout the system. The use of clear headings, bullet points, and visual cues can help students quickly identify important information and navigate through the system more efficiently [28].

Furthermore, the interface should provide students with feedback and progress tracking. This can be done through the use of progress bars, badges, or visual indicators that show students how they are progressing in their learning journey. Feedback can also be provided through personalized messages or notifications that guide students and motivate them to continue their learning [28].

In addition, the user interface should be responsive and accessible across different devices, such as desktop computers, laptops, tablets, and smartphones [29]. This allows students to access an intelligence-based teaching system at the right time and time, providing flexibility and appropriateness to the specific circumstances of the educational process [30].

By designing an effective user interface for the intelligent tutoring system, Iraqi university students can have a seamless and engaging learning experience. A user-friendly interface that is visually appealing, easy to navigate, and provides feedback can greatly enhance the effectiveness of the system and improve students' learning outcomes [31]

2.4 Implementing Adaptive Learning Techniques in the Design Architecture

One of the key features of an intelligent tutoring system is its ability to adapt to the individual needs and learning styles of students. By implementing adaptive learning techniques in the design architecture, Iraqi universities It enables it to provide an educational experience with characteristics that suit students.

By incorporating adaptive learning techniques, the design architecture of the intelligent tutoring system can ensure that students receive the support and guidance they need to succeed. This approach to learning can significantly improve student engagement, motivation, and learning outcomes.

Adaptive learning techniques involve using data and algorithms to analyses the performance and progress of students. This information is then used to adjust the content, difficulty level, and pace of instruction to match the needs of each student. For example, if a student is struggling with a particular concept, the system can provide additional explanations or practice materials to help them grasp the concept better [31].

Furthermore, the use of adaptive learning techniques can also help In determining students' learning mechanisms and methods and their repetitive behaviors. The system can analyses data to identify common misconceptions, areas of weakness, or learning strategies that are effective for specific groups of students. This information can be used to refine and improve the design architecture of the intelligent tutoring system over time [31].

Benefits of Adaptive Learning Techniques:

1. Personalized learning: These technologies provide methods that suit students' individual capabilities and knowledge.
2. Improved Student Engagement: By tailoring the content and instructional approach, adaptive learning techniques can increase student motivation and engagement.
3. Targeted Intervention: The system can identify areas of weakness or misconceptions and provide targeted interventions to help students overcome these challenges.
4. Continuous Improvement: Through data analysis, the design architecture can be refined and improved to Providing an experience built on the students' skills.

Overall, implementing adaptive learning techniques in the design architecture of an intelligent tutoring system can revolutionize the way Iraqi university students learn. By personalizing the learning experience and providing targeted support, these systems have the potential to enhance student outcomes and improve the quality of education in Iraqi universities [32, 33].

3 The Role of Data Analytics in Enhancing the Intelligent Tutoring System

Data analytics plays a crucial role in enhancing the effectiveness of an Intelligent Tutoring System (ITS) for Iraqi university students. The analysis based on ITS leads to evaluating the vision of the educational environment about students' learning patterns in order to modify teaching methods [21, 34].

Data analytics also enables the ITS to track students' progress over time. This information can be used to identify trends and patterns in student performance, helping instructors The educational decision-making strategy is built to develop educational vocabulary based on the actual needs of students, the system can alert instructors to the need for additional support or modifications to the curriculum [28].

Furthermore, data analytics can provide valuable feedback to both students and instructors. Through detailed reports and visualizations, the system can present performance metrics, highlighting areas of strength and weakness. This feedback can guide students in self-assessment and goal-setting, empowering them to take ownership of their learning. Similarly, instructors can use this information to provide targeted feedback and support, fostering a more personalized and effective teaching approach [32, 33].

By utilizing data analytics, the ITS can continuously improve its performance and adapt to the evolving needs of Iraqi university students. The system can leverage machine learning algorithms to analyses data from multiple sources, including student interactions, assessments, and external resources. This analysis can help identify correlations and patterns that may not be immediately apparent to instructors, enabling the system to make more accurate predictions and recommendations [24].

In conclusion, data analytics plays a vital role in enhancing the intelligent tutoring system for Iraqi university students. It enables personalization, tracking progress, providing feedback, and continuous improvement. By harnessing the power of data, the ITS can create a smarter learning environment that caters to the unique needs and challenges of Iraqi students (Fig. 3).

3.1 Addressing Cultural and Language Barriers in the Design Architecture

In order to create a truly effective intelligent tutoring system for Iraqi university students, it is crucial to address the cultural and language barriers that they may face. By understanding and acknowledging these barriers, designers can ensure that the system is accessible and inclusive for all students [35].

1. Cultural Sensitivity: The design architecture should take into account the cultural norms and values of the target audience. For example, in Iraq, there may be certain

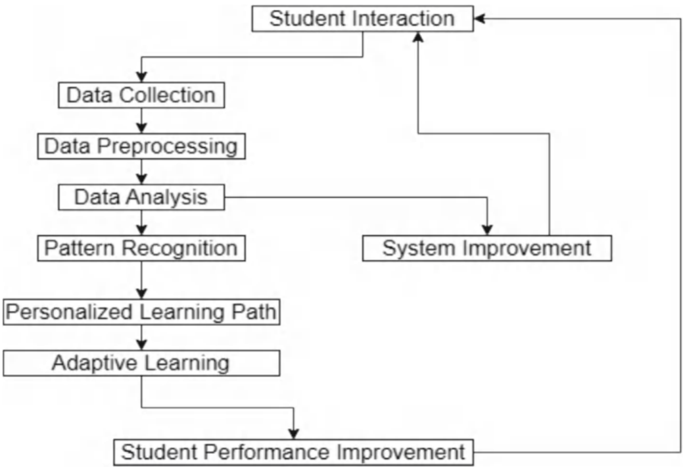


Fig. 3 Data analytics process

cultural sensitivities around topics such as religion or gender. Designers should ensure that the content and examples used in the system are culturally appropriate and do not inadvertently offend or exclude any students [24].

2. Language Support: Iraqi university students may have varying levels of English proficiency. To overcome language barriers, the design architecture should include language support features such as translations, subtitles, or voiceovers in Arabic. This will enable students to fully engage with the system and understand the instructional content without struggling with language comprehension [33].
3. Visual Communication: In addition to language support, the design architecture should prioritize visual communication to aid comprehension for students with different language backgrounds. The use of clear and concise visuals, symbols, and icons can help convey information effectively, even for students who may not be fluent in English [31].
4. Adaptive Learning: The design architecture should also adapt to individual students' cultural and language needs. By incorporating adaptive learning techniques, the system can customize the content and learning experience based on each student's preferences and requirements. For example, the system can offer alternative explanations or examples in the student's native language if they are struggling to understand a concept [31]. These techniques are tailored to individual needs through various methods [36–40]:

- Continuous data collection and analysis.
- Pre-assessments and learning style inventories.
- Student input on goals and preferences.
- Regular reassessment and adjustment.
- Integration of multiple data sources.

1. Collaborative Learning: To further bridge cultural and language barriers, the design architecture Allows for shared learning. By working in groups or pairs, enhancing interaction between students' cultures. This can help students build relationships, develop communication skills, and enhance their understanding of different cultures [35].

By addressing cultural and language barriers in the design architecture, Creating a smart teaching environment that suits Iraqi society more effectively to university students. This will not only improve their academic performance but also promote cultural understanding and diversity within the educational setting [41].

3.2 Integrating Gamification to Improve Student Engagement

Smart education reflects more mature tools in dealing with the university teaching system. It constitutes an important factor that attracts and motivates staff, including students, depending on the type of incentives it provides in facilitating the teaching process in Iraqi universities [42].

One key benefit of gamification is its ability to provide immediate feedback and recognition. For example, when completing a task or answering a question correctly, students can earn virtual points or badges, which can boost their confidence and encourage them to continue learning. The process of motivation gives students a wonderful feeling of continuity and success [30].

In addition, gamification can foster healthy competition among students. By displaying leaderboards that showcase the top performers, the system can create a sense of rivalry and drive students to excel. This competitive element can further enhance student engagement and encourage them to put in extra effort to climb up the rankings [35].

Furthermore, gamification can make learning more immersive and interactive. By incorporating game-like scenarios, simulations, and challenges. The system can create an attractive environment for students based on active, dynamic learning. For example, a language learning module could include interactive exercises where students have to match words with their corresponding images or complete sentences in a timed manner. This interactive approach can make the learning process more enjoyable and memorable [30].

By integrating gamification into the design architecture of an intelligent tutoring system, Iraqi university students can benefit from increased engagement, motivation, and enjoyment in their learning journey. It is necessary to achieve justice among the environment and educational content to ensure that the system remains focused on academic objectives. The gamification features should complement the learning materials and not overshadow them [42].

3.3 Future Prospects and Improvements for Intelligent Tutoring Systems in Iraqi Universities

As technological progress continues, the future prospects for smart teaching systems in Iraqi universities are promising. These systems have the ability to develop methods of learning and acquiring knowledge. With continuous improvements and developments in design engineering, it can provide more personalized and effective support for undergraduate students [42].

Potential Improvements

1. Enhanced Adaptability: Future intelligent tutoring systems can be designed to adapt and cater to the individual learning styles and preferences of Iraqi university students. By analysing data on student performance and engagement. These smart systems work to exploit and support resources to overcome the challenges identified among students [30].
2. Intuitive User Experience: The user interface of intelligent tutoring systems can be further improved to provide a seamless and intuitive learning experience. Clear navigation, simple instructions, and visually appealing interfaces can enhance student engagement and motivation [30].
3. Timing of feedback: It represents a link to integrate smart education systems and provide solutions at the correct time based on students' needs. In addition to the evaluation process based on data analysis for parties to the educational process to enhance confidence in the system's outputs [42], 30.

3.4 Promising Areas

There are several areas where intelligent tutoring systems can make a significant impact on the education system in Iraqi universities:

1. Accessibility: By leveraging technology, intelligent tutoring systems can overcome barriers to education, such as geographical limitations and lack of resources. Iraqi students in remote areas can benefit from quality education and to enable greater resources in the educational environment [41], 43.
2. Personalized Learning: The adaptive nature of intelligent tutoring systems allows for personalized learning experiences. With tailored content and recommendations, students can focus on their individual needs and interests, leading to better learning outcomes [42].
3. Addressing Language Barriers: Iraqi universities have diverse student populations, often with differing levels of English proficiency. Intelligent tutoring systems can provide language support and translation features to ensure that language barriers do not hinder students' learning experiences [32].

Overall, the future of intelligent tutoring systems in Iraqi universities looks promising. With continuous advancements in design architecture and the incorporation of adaptive learning techniques and data analytics. Construction is often done to include smart technology that creates a comprehensive change in the university environment and enables students to achieve their goals..

4 Conclusion

The design architecture of the education system aims to enhance the capabilities of universities and understand the challenges that hinder students. Through their distinct components such as adaptive learning and data analytics, these systems can provide effective self-support.

The user interface plays a crucial role in the success of an intelligent tutoring system, as it should be intuitive, visually appealing, and easy to navigate. Additionally, addressing cultural and language barriers is essential to ensure that the system is accessible and inclusive for all students.

The future opens many horizons for improving smart teaching systems. By integrating new technologies such as artificial intelligence and virtual reality, which provides more efficient and interactive educational experiences. Continuing research on developing education systems will contribute to advancing innovation in this field.

Overall, the design architecture of an intelligent tutoring system can revolutionize education in Iraqi universities by providing personalized and adaptive support to students, ultimately enhancing their learning outcomes and academic success.

References

1. Abdulwahid, H. S., Salman, M. K., & Ouda, A. J. (2019). Identify and assessing the e-readiness of public organizations (case study in Iraq). Paper presented at the AIP Conference Proceedings.
2. Ahmed, K. A., Mansoor, M. S., Al-Imareen, N., & Alameri, I. (2022). Evaluation of E-Teaching Implementation in Iraqi Universities. Paper presented at the Congress on Intelligent Systems.
3. Alkhuwaylidee, A. R. (2019). Extended unified theory acceptance and use technology (utaut) for e-learning. Journal of Computational and Theoretical Nanoscience, 16(3), 845-852.
4. Crow, T., Luxton-Reilly, A., & Wuensche, B. (2018). Intelligent tutoring systems for programming education: a systematic review. Paper presented at the Proceedings of the 20th Australasian Computing Education Conference.
5. Albornoz-De Luise, R. S., Arevalillo-Herráez, M., & Arnau, D. (2023). On using conversational frameworks to support natural language interaction in intelligent tutoring systems. IEEE Transactions on Learning Technologies, 16(5), 722-735.
6. Hemachandran, K., Verma, P., Pareek, P., Arora, N., Rajesh Kumar, K. V., Ahanger, T. A., . . . Ratna, R. (2022). Artificial intelligence: A universal virtual tool to augment tutoring in higher education. Computational Intelligence and Neuroscience, 2022(1), 1410448.
7. Alam, A. (2022). Employing adaptive learning and intelligent tutoring robots for virtual classrooms and smart campuses: reforming education in the age of artificial intelligence Advanced computing and intelligent technologies: Proceedings of ICACIT 2022 (pp. 395–406): Springer.
8. González-Calatayud, V., Prendes-Espinosa, P., & Roig-Vila, R. (2021). Artificial intelligence for student assessment: A systematic review. Applied Sciences, 11(12), 5467.
9. Hsieh, M.-C., & Chen, S.-H. (2019). Intelligence augmented reality tutoring system for mathematics teaching and learning. Journal of Internet Technology, 20(5), 1673-1681.

10. Yuce, A., Abubakar, A. M., & Ilkan, M. (2019). Intelligent tutoring systems and learning performance: Applying task-technology fit and IS success model. Online Information Review, 43(4), 600-616.
11. Gardner, J., O'Leary, M., & Yuan, L. (2021). Artificial intelligence in educational assessment:'Breakthrough? Or buncombe and ballyhoo?'. Journal of Computer Assisted Learning, 37(5), 1207-1216.
12. Oh, C., Song, J., Choi, J., Kim, S., Lee, S., & Suh, B. (2018). I lead, you help but only with enough details: Understanding user experience of co-creation with artificial intelligence. Paper presented at the Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.
13. Wang, Z., Yan, W., Zeng, C., Tian, Y., & Dong, S. (2023). A unified interpretable intelligent learning diagnosis framework for learning performance prediction in intelligent tutoring systems. International Journal of Intelligent Systems, 2023(1), 4468025.
14. Chen, X., Zou, D., Xie, H., Cheng, G., & Liu, C. (2022). Two decades of artificial intelligence in education. Educational Technology & Society, 25(1), 28-47.
15. Salem, A.-B. M., Mikhalkina, E. V., & Nikitaeva, A. Y. (2019). Establishment of smart education system in modern universities: Concept, technologies and challenges. International journal of education and information technologies, 13, 180-188.
16. Alrakhawi, H. A., Jamiat, N., & Abu-Naser, S. S. (2023). Intelligent Tutoring Systems in education: A systematic review of usage, tools, effects and evaluation. Journal of Theoretical and Applied Information Technology, 101(4), 1205-1226.
17. Woolf, B. P. (2010). Building intelligent interactive tutors: Student-centered strategies for revolutionizing e-learning: Morgan Kaufmann.
18. Penalva, J. (2019). Contemporary methodological perspectives in educational research on 'teachers' practice': Assumptions and shortcomings for 'effective practices'. European Educational Research Journal, 18(4), 461-482.
19. SHARP, L., Carruba-Rogel, Z., & Diego-Medrano, E. (2019). Strengths and shortcomings of a teacher preparation program: Learning from racially diverse preservice teachers. Journal of Teacher Education and Educators, 8(3), 281-301.
20. Mohamed, H., & Lamia, M. (2018). Implementing flipped classroom that used an intelligent tutoring system into learning process. Computers & Education, 124, 62-76.
21. Ge, L., Chen, Y., Yan, C., Chen, Z., & Liu, J. (2020). Effectiveness of flipped classroom vs traditional lectures in radiology education: A meta-analysis. Medicine, 99(40), e22430.
22. Nhongo, R., & Tshotsho, B. P. (2021). The shortcomings of emergency remote teaching in rural settings of Zimbabwe during COVID-19 school closures: Lessons from China's experience. Africa's Public Service Delivery & Performance Review, 9(1), 9.
23. French, R., Imms, W., & Mahat, M. (2020). Case studies on the transition from traditional classrooms to innovative learning environments: Emerging strategies for success. Improving Schools, 23(2), 175-189.
24. Mousavinasab, E., Zarifsanaiey, N., R. Niakan Kalhori, S., Rakhshan, M., Keikha, L., & Ghazi Saeedi, M. (2021). Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments, 29(1), 142-163.
25. Heilporn, G., Lakhal, S., & Bélisle, M. (2021). An examination of teachers' strategies to foster student engagement in blended learning in higher education. International Journal of Educational Technology in Higher Education, 18(1), 25.
26. Kahu, E. R., & Nelson, K. (2018). Student engagement in the educational interface: Understanding the mechanisms of student success. Higher education research & development, 37(1), 58-71.
27. Bond, M., Buntins, K., Bedenlier, S., Zawacki-Richter, O., & Kerres, M. (2020). Mapping research in student engagement and educational technology in higher education: A systematic evidence map. International Journal of Educational Technology in Higher Education, 17, 1-30.
28. Tight, M. (2020). Student retention and engagement in higher education. Journal of further and Higher Education, 44(5), 689-704.

29. Khalaf Gatea, Ali, Anwar Adnan Alnawas, and Zina Hamza Gali. 2021. The Effect of XBRL Financial Reporting on Enhancing the Transparency of Information in the Financial Statements. Vol. 12.
30. Paladines, J., & Ramirez, J. (2020). A systematic literature review of intelligent tutoring systems with dialogue in natural language. IEEE Access, 8, 164246-164267.
31. Alam, A. (2023). Harnessing the Power of AI to Create Intelligent Tutoring Systems for Enhanced Classroom Experience and Improved Learning Outcomes Intelligent Communication Technologies and Virtual Mobile Networks (pp. 571–591): Springer.
32. Azevedo, R., Bouchet, F., Duffy, M., Harley, J., Taub, M., Trevors, G., . . . Wortha, F. (2022). Lessons learned and future directions of metatutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. Frontiers in Psychology, 13, 813632.
33. Mathan, S. A., & Koedinger, K. R. (2018). Fostering the intelligent novice: Learning from errors with metacognitive tutoring Computers as Metacognitive Tools for Enhancing Learning (pp. 257–265): Routledge.
34. Peters, H., Zdravkovic, M., João Costa, M., Celenza, A., Ghias, K., Klamen, D., … Wangsaturaka, D. (2019). Twelve tips for enhancing student engagement. Medical teacher, 41(6), 632-637.
35. Alkhatlan, A., & Kalita, J. (2018). Intelligent tutoring systems: A comprehensive historical survey with recent developments. arXiv preprint arXiv:1812.09628.
36. Andrade, Heidi L. 2019. "A Critical Review of Research on Student Self-Assessment." Frontiers in Education 4. https://doi.org/10.3389/feduc.2019.00087.
37. Broadbent, J., and W. L. Poon. 2015. "Self-Regulated Learning Strategies & Academic Achievement in Online Higher Education Learning Environments: A Systematic Review." The Internet and Higher Education 27:1–13. https://doi.org/10.1016/j.iheduc.2015.04.007.
38. Ferguson, Rebecca, and Doug Clow. 2017. "Where Is the Evidence?" Pp. 56–65 in Proceedings of the Seventh International Learning Analytics & Knowledge Conference. New York, NY, USA: ACM.
39. Kirschner, Paul A. 2017. "Stop Propagating the Learning Styles Myth." Computers & Education 106:166–71. https://doi.org/10.1016/j.compedu.2016.12.006.
40. Siemens, George. 2013. "Learning Analytics." American Behavioral Scientist 57(10):1380–1400. https://doi.org/10.1177/0002764213498851.
41. Al Rekhawi, H. A., & Abu Naser, S. S. (2018). An intelligent tutoring system for learning Android applications UI development. International Journal of Engineering and Information Systems (IJEAIS), 2(1), 1-14.
42. Dermeval, D., Paiva, R., Bittencourt, I. I., Vassileva, J., & Borges, D. (2018). Authoring tools for designing intelligent tutoring systems: a systematic review of the literature. International Journal of Artificial Intelligence in Education, 28, 336-384.
43. VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. Educational psychologist, 46(4), 197-221.
44. Hasan, R. A., Alomari, M. F., & Jamaluddin, J. B. (2023, July). Comparative study: Using machine learning techniques about rainfall prediction. In AIP Conference Proceedings (Vol. 2787, No. 1). AIP Publishing.

Automated Mandible Segmentation from Computed Tomography Scans Using U-Net and U-Net Transformer

Robert William Wacan, Louisse Amadeo Romero, Paul Justine Ardiente, Lemuel Clark Velasco, Mohana Shanmugam, and Chris Jordan Aliac

Abstract Automated segmentation of mandibular bones in computed tomography (CT) scans is crucial for craniomandibular surgeries. Since traditional manual or semi-automated methods face challenges due to labor intensiveness and interobserver variability, deep learning, notably the U-Net architecture, offers a promising solution by accurately delineating anatomical structures in CT scans. This study compared a traditional 2D U-Net and U-Net Transformer model in implementing automated mandible segmentation of head and neck CT scans. The U-Net transformer model exhibited slightly higher Dice similarity coefficient scores and Intersection over Union (IoU) values compared to the U-Net model, indicating improved segmentation efficiency and spatial overlap with ground truth annotations. Additionally, the U-Net Transformer model also achieved superior pixel classification accuracy compared to the 2D U-Net model. These findings suggest that the U-Net Transformer model can significantly enhance surgical planning and diagnosis in craniofacial surgeries, highlighting the potential of Transformer-based deep learning architectures in automating mandibular segmentation from CT scans.

R. W. Wacan · L. A. Romero · P. J. Ardiente · C. J. Aliac (✉)
Department of Computer Science, College of Computer Studies, Cebu Institute of Technology University, Cebu, The Philippines
e-mail: chris.aliac@cit.edu

C. J. Aliac
Artificial Intelligence Fablab, Cebu Institute of Technology University, Cebu, The Philippines

L. C. Velasco
Department of Information Technology, College of Computer Studies, Mindanao State University-Iligan Institute of Technology, Iligan, The Philippines

Center for Computational Analytics and Modelling, Premier Research Institute of Science and Mathematics, Mindanao State University-Iligan Institute of Technology, Iligan, The Philippines

Department of Industrial and Information Management, College of Management, National Cheng Kung University, Tainan City, Taiwan

M. Shanmugam
Department of Informatics, College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Putrajaya, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_6

Keywords Automated image segmentation · Mandibular bone · Computed tomography scans · Deep learning methods · U-net first section

1 Introduction

Image segmentation is pivotal in various medical fields, including dentistry and oral surgery, where accurate delineation of anatomical structures like the mandibular bone from Computed Tomography (CT) scans is essential. The process faces numerous challenges due to the intricate anatomy of the head and neck, the presence of metal artifacts, and the variability in bone structures. Manual segmentation, though accurate, is impractical for widespread clinical use due to its time-consuming nature and lack of reproducibility [1, 2]. The usage of CT scan files has the potential to significantly improved the diagnosis and classification of facial conditions which could assist clinicians and surgeons with assessments and personalized treatment plans. Automated mandible segmentation has emerged as a crucial trend, driven by the need to reduce reliance on expert knowledge and overcome the limitations of manual segmentation [2–4]. Addressing the challenges in mandible segmentation is crucial for advancing clinical applications such as surgical planning and radiotherapy. Moreover, the complexity of head and neck CT scans, which include various bone structures and artifacts from metal implants, further complicates the process [3, 5]. In the context of automated mandible segmentation from CT scans, deep learning methods have shown substantial potential. Deep learning methods have already revolutionized various aspects of medical image analysis. These methods leverage large datasets and complex algorithms to learn intricate patterns and features, significantly improving the accuracy and reliability of segmentation tasks [1, 2, 4]. Specifically, the U-Net architecture has been widely adopted for its ability to accurately segment medical images across different modalities. Recent advancements such as Focus U-Net, 3D U-Net, and UNet++ have further enhanced its capabilities [3, 6–9]. Focus U-Net, a novel dual attention-gated CNN, has emerged as a powerful tool for polyp segmentation during colonoscopy, significantly advancing colorectal cancer screening [8, 10]. UNet++ , a nested U-Net architecture, has demonstrated exceptional performance in medical image segmentation tasks, offering accurate and detailed anatomical delineation [7, 8, 11]. Moreover, the integration of transformer with U-Net is still an underexplored yet promising field of study in the applications of deep learning for medical imaging [12–14]. These advancements if applied properly can efficiently assist clinicians and surgeons in mandible segmentation.

Accurately segmenting the mandible with CT scans as inputs can be challenging but can be achievable with proper deep learning implementation. The problem in properly implementing the appropriate deep learning intervention lies in the complexity of the scans and the scarcity of experienced professionals that links both the technical and surgical field [4, 5, 7, 10]. The current manual segmentation methods often rely heavily on expert knowledge of the clinicians and surgeons, posing challenges due to a shortage of these experienced professionals. In order to address this problem, the objective of this study is to evaluate and compare the effectiveness of two deep learning methods—traditional U-Net and the innovative U-Net Transformer—in automating mandibular segmentation in CT scans that can potentially assist clinicians and surgeons in their clinical decisions. Deep learning-based techniques have the potential to solve the shortage of expert professionals while at the same time achieving remarkable accuracy in surgical planning by offering precise intervention suggestions [2–4, 11]. By examining U-Net and U-Net-Transformer implementations on the same dataset, this study hopes to highlight their strengths and limitations, providing insights into which approach offers more accurate delineation of the mandibular bone. This study aims to build on these advancements by exploring the potential of deep learning methods to further improve segmentation accuracy and efficiency, ultimately enhancing patient care and outcomes in oral and maxillofacial surgery.

This paper begins by discussing the significance of automated mandible segmentation and the role of deep learning models in enhancing medical imaging techniques. It details the methodology, including data preparation from 48 DICOM files, manual segmentation, and partitioning into training, validation, and testing sets. The U-Net and U-Net Transformer architectures are described, highlighting their design and training processes, along with evaluation using metrics like the Dice similarity coefficient, IoU, and accuracy. The results section compares the models, with the U-Net Transformer showing superior performance in segmentation accuracy, supported by visual comparisons and metric analysis. The paper concludes by emphasizing the U-Net Transformer's potential for clinical applications and suggests future research directions, including exploring multimodal data fusion, enhancing dataset diversity, and utilizing transfer learning for improved segmentation performance.

2 Methodology

2.1 CT Scan Data Preparation

This study focused specifically on assessing the viability of training using a smaller dataset of only 48 DICOM files [6, 8, 11, 15, 16]. An open-source, anonymized Head and Neck CT scan dataset was sourced online from the "Mandibular CT Dataset Collection" available on the website called "Figshare" authored by Jürgen Wallner and Jan Egger, comprising 48 different CT scans in DICOM format, the standard for medical images [17, 18]. These scans were manually segmented to create ground truth masks, ensuring high-quality annotations essential for training robust models [3, 4, 19]. The dataset was then randomly partitioned into training set with 36 CT scans, validation set with 6 CT scans, and testing set with 6 CT scans to maintain patient-level independence across subsets. Preprocessing steps were then conducted which includes resizing and normalization as well as standardization of the resolution and quality of all CT scans [10, 11, 20]. The annotated data was organized into specific folders: mandible-segmentation, masks, train images, train masks, val images, and val masks, ensuring a structured approach to model training and validation. Figures 1, 2, and 3 illustrate the progression from raw CT scans to manually segmented mandibles, and finally to detailed segmented mandibular structures in the axial view, highlighting the accuracy and consistency of the segmentation across multiple slices. Figure 1 shows that this raw CT scan offers an unfiltered view of the scanned area serving as the foundation of our analysis, providing a starting point for understanding the complexities inherent in the imaging data.

By presenting the raw scan in its original form, the researchers aim to convey the intricacies of the anatomical structures captured within the image, including bones, soft tissues, and other relevant features. The presentation of the raw CT scan prompts considerations regarding image quality, resolution, and noise levels, all of which can impact the accuracy and reliability of subsequent processing steps [1, 21, 22]. Figure 2 further shows the segmentation step which involves a detailed and careful delineation of the mandibular structures from surrounding tissues, aiming to isolate the mandible for further analysis.

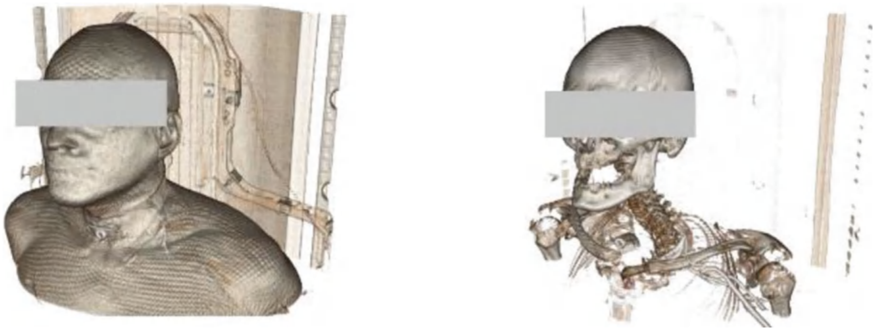


Fig. 1 One of raw and anonymized head and neck CT scans in the study's dataset

Fig. 2 One of the manually segmented mandibles from a head and neck CT Scan

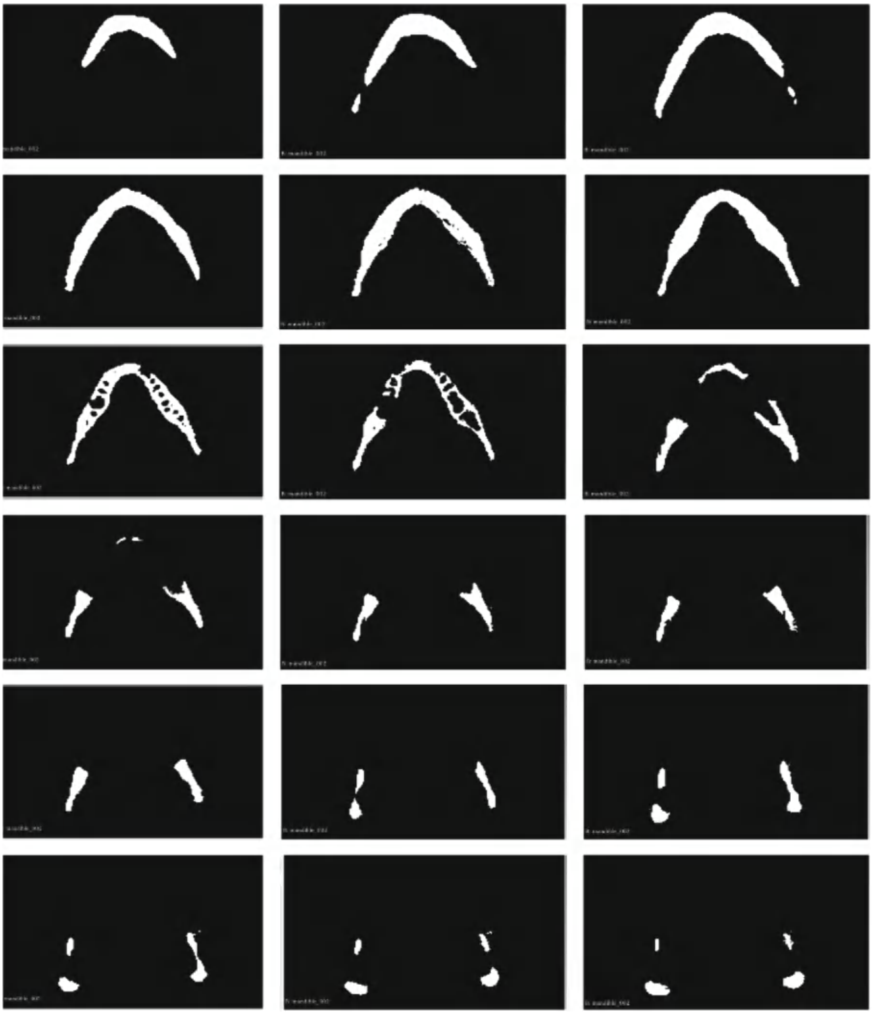


Fig. 3 Segmented mandibular bone and teeth on each DICOM slice—axial view

Through manual tracing and outlining, the boundaries of the mandibular bone are precisely defined, ensuring accurate representation and separation from adjacent anatomical elements. By focusing on this manual segmentation process, the researchers addressed the challenges and considerations involved in accurately delineating anatomical structures within medical imaging data [5, 20]. Additionally, this data preprocessing addressed the importance of manual segmentation in refining and preparing the data for subsequent quantitative analysis, underscoring its role in enhancing the accuracy and reliability of research findings. Figure 3 then shows the three-dimensional reconstruction of the mandible derived from the segmented CT scan data after the manual segmentation process.

This advanced visualization technique offers a comprehensive representation of the mandibular anatomy in three-dimensional space, providing insights into its spatial relationships and structural integrity. Through this 3D reconstruction, the researchers addressed the transformative potential of advanced imaging technologies in medical research, particularly in the enhanced understanding of the complex anatomical structures. The conducted comprehensive data preparation process which involved data selection, cleaning, annotation, and partitioning was considered an important requisite before model implementation and evaluation.

2.2 Model Implementation and Evaluation

This study focuses on the comparison of the implementation of a U-Net architecture and U-Net Transformer architecture for automating mandible segmentation. Figure 4 shows that the architecture of the U-Net model was outlined, and necessary modifications or enhancements were made as required. The 2D U-Net involves a contracting path for downsampling and an expansive path for upsampling, using convolutional layers, batch normalization, and ReLU activations. It features skip connections for detail retention and transposed convolutions for refining spatial features. Implemented in PyTorch, the model supports both 2D and 3D segmentation. The model training process was then executed in two distinct phases. Initially, model training commenced using segmented data to establish an initial model. Subsequently, fine-tuning was carried out using non-segmented data, allowing for the iterative refinement of the model to achieve optimal segmentation performance.

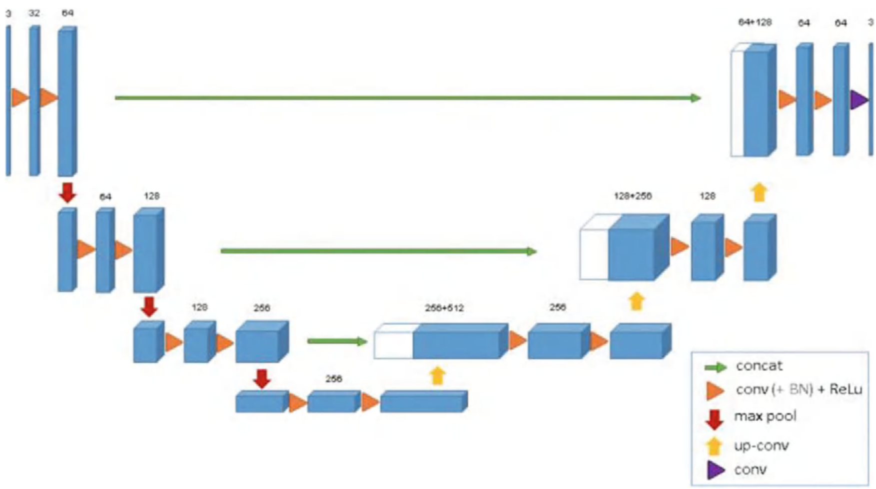


Fig. 4. 2D U-Net architecture as implemented in this study

The figure above illustrates the U-Net architecture, a convolutional neural network designed for image segmentation. The diagram breaks down the network into its key components, highlighting the flow of data through the network. Each part of the figure is explained by the ff:

Input Layer: The input image is represented by the blue rectangles at the top left. The input dimensions are 32 × 32 pixels with 64 channels (which could represent features or different input images).

First Block: Convolutional Layer + Batch Normalization (BN) + ReLU: The orange triangles indicate convolutional layers followed by batch normalization and ReLU activation. The numbers beside these layers (e.g., 64) denote the number of feature channels.

Max Pooling: The orange triangles pointing down represent max pooling operations that reduce the spatial dimensions by half (e.g., from 32 × 32 to 16 × 16).

Subsequent Blocks: This process of applying convolutional layers followed by max pooling is repeated, each time doubling the number of feature channels and halving the spatial dimensions, until the bottleneck layer is reached. For example: from 32 × 32 with 64 channels to 16 × 16 with 128 channels, and so on. Each block outputs feature maps with increasingly abstract representations of the input image.

Middle of the "U" model process: The bottleneck layer at the bottom of the "U" has the smallest spatial dimensions but the highest number of feature channels. This layer captures the most abstract features of the input image. Here, convolutional operations are applied without any further max pooling.

Up-Convolution (Transposed Convolution): The yellow triangles pointing up represent up-convolutions that increase the spatial dimensions (e.g., from 8 × 8 to 16 × 16) and halve the number of feature channels.

Concatenation: The green arrows indicate concatenation of the upsampled feature maps with the corresponding feature maps from the contracting path. This operation combines the high-resolution features from the contracting path with the upsampled features, helping the network to retain spatial information.

Continuous Expanding blocks: This process of up-convolution followed by concatenation and convolution is repeated, each time increasing the spatial dimensions and reducing the number of feature channels until the original input dimensions are restored.

Output Layer: The output of the last expanding block is a set of feature maps with the same spatial dimensions as the input image.

Convolutional Layer: The purple triangles represent the final convolutional layer that reduces the number of feature channels to the desired number of output classes. For binary segmentation, this would typically be a single channel with a sigmoid activation function to predict the probability of each pixel belonging to the target class.

Similar to the approach taken with the U-Net model, the U-Net Transformer architecture was carefully outlined, and any requisite modifications or enhancements were made. Figure 5 shows that the U-Net Transformer architecture introduces a hierarchical encoder-decoder structure. In this approach, 3D patches are fed into the encoder, and its outputs are connected to the decoder through skip connections. This design allows U-Net Transformer to capture both local and global contextual information effectively. The U-Net Transformer uses batch normalization for stable training, complemented by transposed convolutions for refining upsampling. While the traditional U-Net architecture is primarily designed for 2D image segmentation tasks, U-Net Transformer is adapted to handle 3D data more effectively. This adaptation makes U-Net Transformer suitable for volumetric medical image segmentation tasks, where analyzing information across multiple dimensions where depth, width, and height is considered crucial. U-Net Transformer was implemented in TensorFlow to accommodates both 2D and 3D data. The training process adopted a two-phase strategy with an initial model establishment through segmented data training, followed by iterative refinement via fine-tuning with non-segmented data, aimed at optimizing segmentation performance. An early stopping strategy was utilized if there was no improvement in the loss of the validation set.

Part of the U-Net Transformer is an advanced neural network architecture designed specifically for image segmentation tasks, known for its ability to produce precise and reliable segmentations by effectively capturing both high-level semantic information and fine-grained details. The architecture of U-Net Transformer likely comprises a contracting path or encoder to capture context and a symmetric expanding path or decoder to enable precise localization. This dual-path structure allows the network to learn a rich hierarchy of features, facilitating accurate segmentation of complex structures within the images. The encoder path reduces the spatial dimensions through pooling operations while increasing the depth of feature maps, which captures the broader context. Conversely, the decoder path performs upsampling and concatenates

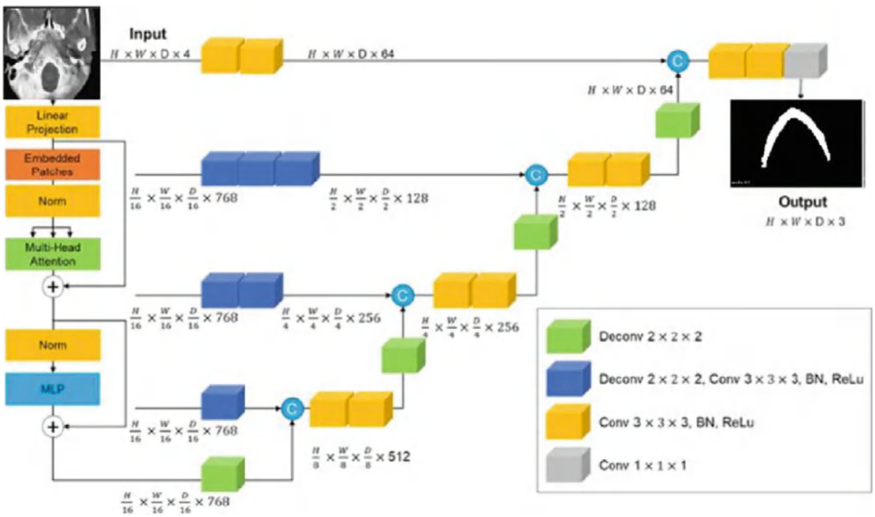


Fig. 5 U-Net Transformer architecture as implemented in this study

these high-level feature maps with corresponding feature maps from the encoder path, ensuring the restoration of spatial information and fine details.

Since the core of this research involved a detailed comparison of the implementation of a 2D U-Net architecture and UNet-R architecture for automating mandible segmentation, the 2D U-Net architecture was outlined, involving a contracting path for downsampling and an expansive path for upsampling, with skip connections for detail retention and transposed convolutions for refining spatial features. On the other hand, the UNet-R architecture introduced a hierarchical encoder-decoder structure, capturing both local and global contextual information effectively. Both architectures underwent a two-phase training process: initial model establishment using segmented data and iterative refinement via fine-tuning with non-segmented data [6, 19]. Since the focus shifts to evaluating segmentation performance using three key overlap-based metrics: The Dice similarity coefficient, Intersection over Union (IoU), and accuracy. These metrics are essential for gauging the alignment between algorithmic segmentation and manual reference, providing insights into the degree of overlap and overall segmentation accuracy. Equation 1 shows that the Dice similarity coefficient is a primary measure for assessing the similarity between two sets by quantifying their overlap with TP as the true positive, FP as the false positive and FN as the false negative. The Dice coefficient is calculated by doubling the intersection of the two sets and dividing it by the sum of their sizes. This metric is fundamental in comparing the segmented region to the ground truth, offering a concise evaluation of segmentation accuracy [9, 23]. Expected results ranges from 0 to 1, where 0 indicates no overlap and 1 signifies perfect agreement.

$$Dice = \frac{2TP}{FP + TP + FN} \tag{1}$$

IoU, also known as the Jaccard index, is another crucial metric used in the study since it quantifies the degree of overlap by assessing the ratio of the intersection to the union of the predicted and reference regions. Equation 2 where area of overlap is divided by the area of union shows that IoU provides a valuable measure of segmentation accuracy by indicating how well the algorithmically generated segmentation matches the ground truth [23, 24].

$$IOU = \frac{A \cap B}{A \cup B} \tag{2}$$

Accuracy is then used by the researchers as an overlap-based metric in the evaluation phase. Accuracy measures the proportion of true positive classifications, indicating the number of pixels or voxels correctly identified as part of the target object by both the segmentation algorithm and the reference standard [2, 19, 25]. This metric was used by the researchers to reflect the model's capability to accurately identify foreground elements within the image data, thus offering a comprehensive perspective on segmentation quality.

2.3 Deep Learning Methods Versus Other Studies

In this paper, the mandible segmentation approach uses two Artificial Intelligent deep learning neural network model (DLNN). The advantage of a Deep learning neural network is the ability of taking in raw, non-pre-processed images as training data. It has the ability of filtering out unwanted sections of the image that are not needed in the segmentation process [1, 2, 6–10]. Other studies use Principal Component analysis (PCA) in the construction templates that identifies the shape of the objects [16–18]. Several papers use statistical shape model-based approaches (SSMB) deployed over simple neural network models and or statistical based static AI algorithms [20, 22, 23]. In this paper, with the use of DLNN, regions of interest (ROI) are automatically segmented in MRI image slices. This approach method maybe depicted as an automated version of the manual segmentation process [24, 25]. MRI data used in the training process was fully annotated to indicate full segmentation, thus, the application training style of the models follows a fully supervised training manner [26, 27].

3 Result and Discussion

The results indicate a clear performance difference between the U-Net and the U-Net Transformer model in terms of segmentation accuracy, as illustrated in Table 1. The Dice score for the U-Net after 40 epochs was 0.504, which points to a relatively lower level of overlap or similarity between the predicted segmentation and the ground truth. This score reflects that the U-Net, with its shorter training period, struggled to achieve high accuracy in segmenting the images [3, 6, 9, 23]. Conversely, the U-Net Transformer model, which was trained for 118 epochs utilizing an early stopping strategy to prevent overfitting, achieved a higher Dice score of 0.554. This improvement suggests that the additional training epochs and the architecture of U-Net Transformer allowed it to better capture the relevant features in the data, resulting in more accurate segmentations [12–14, 23]. The moderate improvement in the Dice score from 0.504 for the U-Net to 0.554 for the U-Net Transformer model, indicates that while the model showed a notable enhancement in performance, there remains room for further optimization and improvement. The early stopping strategy was effective in determining the optimal training duration, preventing the model from overfitting while ensuring that it trained sufficiently to capture the necessary details for accurate segmentation. These results underscore the importance of model architecture and training strategies in achieving high-quality segmentation outcomes. Future work could explore additional techniques to further enhance the Dice score, such as data augmentation, more advanced regularization methods, or further refinements to the U-Net Transformer model. Comparison of the 2 models show the following key differences:

Table 1 Comparison of dice similarity coefficient

Model	Epochs	Dice score
U-Net	40	0.504
U-Net transformer	118	0.554

Table 2 Comparison of intersection over union

Model	Epochs	IoU
U- Net	40	0.558
U-Net transformer	118	0.569

IoU measurements were conducted by comparing the pixel-wise overlap between the predicted segmentation masks and the manually segmented masks. This metric provides a precise quantification of each model's segmentation accuracy. For the U-Net model, the IoU was measured at 0.558, while the U-Net Transformer model achieved a slightly higher IoU of 0.569, as shown in Table 2. These IoU values indicate that both models attained moderate levels of overlap or similarity between their segmentation outputs and the ground truth annotations. The marginally higher IoU value for the U-Net Transformer model suggests a slight performance advantage in capturing the spatial overlap between the predicted and ground truth segmentations [6, 8, 14, 24]. However, the small difference in IoU between the two models highlights that there remains significant room for improvement in segmentation accuracy for both models.

The U-Net Transformer outperforms the U-Net, achieving a higher Dice score (0.554 vs. 0.504). This improvement in performance can be attributed to the U-Net Transformer's ability to capture long-range dependencies and global context through its self-attention mechanisms.

The relatively small increase in the Dice score (0.05) despite the significantly higher number of epochs suggests that while the U-Net Transformer does offer improved performance, the gains might come with diminishing returns in terms of computational resources and training time.

3.1 Differences Between the AI Models

Comparison of the 2 models show the following key differences:

Mechanism of Feature Extraction:

U-Net: Relies primarily on convolutional operations and max pooling to extract and downsample features.

U-Net Transformer: Incorporates Transformer blocks with self-attention mechanisms to capture both local and global features.

Handling Spatial Relationships:

U-Net: Uses skip connections to combine features from different layers, maintaining spatial relationships through convolutions.

U-Net Transformer: Uses self-attention to capture spatial relationships, potentially offering better handling of long-range dependencies.

Complexity and Computational Requirements:

U-Net: Generally simpler and computationally less intensive compared to Transformer-based models.

U-Net Transformer: More complex and computationally demanding due to the inclusion of Transformer blocks, requiring more memory and processing power.

Performance on Large-Scale and Complex Data:

- **U-Net:** Effective for smaller datasets and tasks where local context is sufficient.
- **U-Net Transformer:** Superior for large-scale and complex datasets where understanding global context and long-range dependencies is crucial.

Model Architectural Performance Based on Application of Study:

U-net: Simpler and faster to train, effective with limited data, and achieves moderate performance quickly but might struggle with capturing global context and long-range dependencies, which can limit its performance on more complex segmentation tasks.

U-Net Transformer: Superior performance in capturing global context and long-range dependencies, leading to better segmentation accuracy but more complex and computationally intensive, requiring significantly more training epochs and resources to reach its full potential.

If training time and computational resources are limited, U-Net may be preferable due to its faster convergence. However, if higher segmentation accuracy is paramount and resources allow, the U-Net Transformer offers better performance despite its longer training time.

Sample Heading (Third Level). Only two levels of headings should be numbered. Lower level headings remain unnumbered; they are formatted as run-in headings.

Sample Heading (Forth Level). The contribution should contain no more than four levels of headings. The following **Error! Reference source not found.** gives a summary of all heading levels.

3.2 Accuracy Results on the 2 AI Models

Accuracy was assessed using the true positives metric, which measures the number of pixels or voxels correctly classified as part of the object of interest by both the segmentation algorithm and the ground truth annotation. The U-Net model achieved an accuracy of 99.17%, indicating a high level of correctness in pixel classification, while the U-Net Transformer model exhibited an even higher accuracy of 99.72%, as detailed in Table 3. These high accuracy rates demonstrate that both models are highly effective at correctly identifying the foreground elements within the image data, with the U-Net Transformer model showing superior performance.

Table 3 Comparison of accuracy

Model	Epochs	Accuracy (%)
U- Net	40	99.17
U-Net transformer	118	99.72

While, visual comparisons of the segmentation outputs from the U-Net and U-Net Transformer models underscore their performance differences, both models demonstrate remarkable consistency in segmenting the lower mandible, a region where the anatomy is consistent across scans. Figure 6 illustrates that the minor differences for slices number 57–67 do not significantly affect the overall delineation of the mandible, underscoring the reliability of both models in capturing detailed anatomical features. These findings highlight the models' efficacy and robustness, suggesting areas of stability within the segmentation algorithms that can be leveraged for future refinements to enhance accuracy and reliability. However, a closer examination of the middle part of the mandible, as shown in slice number 69–79 reveals a performance gap between the U-Net and the U-Net Transformer models. The U-Net struggles with accurately segmenting the upper parts of the mandible, showing more inconsistencies and errors, particularly in capturing fine details. In contrast, the U-Net Transformer model demonstrates greater accuracy and detail in its segmentation predictions, highlighting its superior capability in handling complex anatomical features. In the upper part of the mandible, shown in slice number 81–87, the U-Net model exhibits numerous false positives, incorrectly identifying non-mandibular structures as part of the mandible. This misclassification leads to distorted segmentation outputs and potentially inaccurate clinical interpretations. Conversely, the U-Net Transformer model, with fewer false positives, shows more precise delineation of mandibular boundaries, suggesting its greater utility for surgical planning and diagnosis in craniofacial surgeries. The improved accuracy of the U-Net Transformer model indicates its potential for more accurate and reliable clinical applications.

The results of this study on mandible segmentation using the U-Net model compared to the U-Net Transformer model align with the broader advancements in deep learning for medical image analysis. Previous research has demonstrated the transformative impact of deep learning methodologies in segmenting intricate materials and bioimages [6, 12, 14, 26]. Similarly, the results of this study shows an improvement in segmentation accuracy with the U-Net Transformer model, indicating the effectiveness of advanced deep learning architectures in achieving better segmentation outcomes. In the context of medical image analysis, deep learning algorithms have significantly improved disease diagnosis and treatment planning [19–21, 25]. The results of this study, particularly the higher accuracy and better segmentation performance of the U-Net Transformer model, are consistent with these advancements, highlighting the potential of neural networks in enhancing medical imaging processes. Furthermore, the success of U-Net architectures and their variants in medical image segmentation has been widely recognized [3, 6, 8, 9]. Our use of the U-Net Transformer model, which demonstrated superior performance

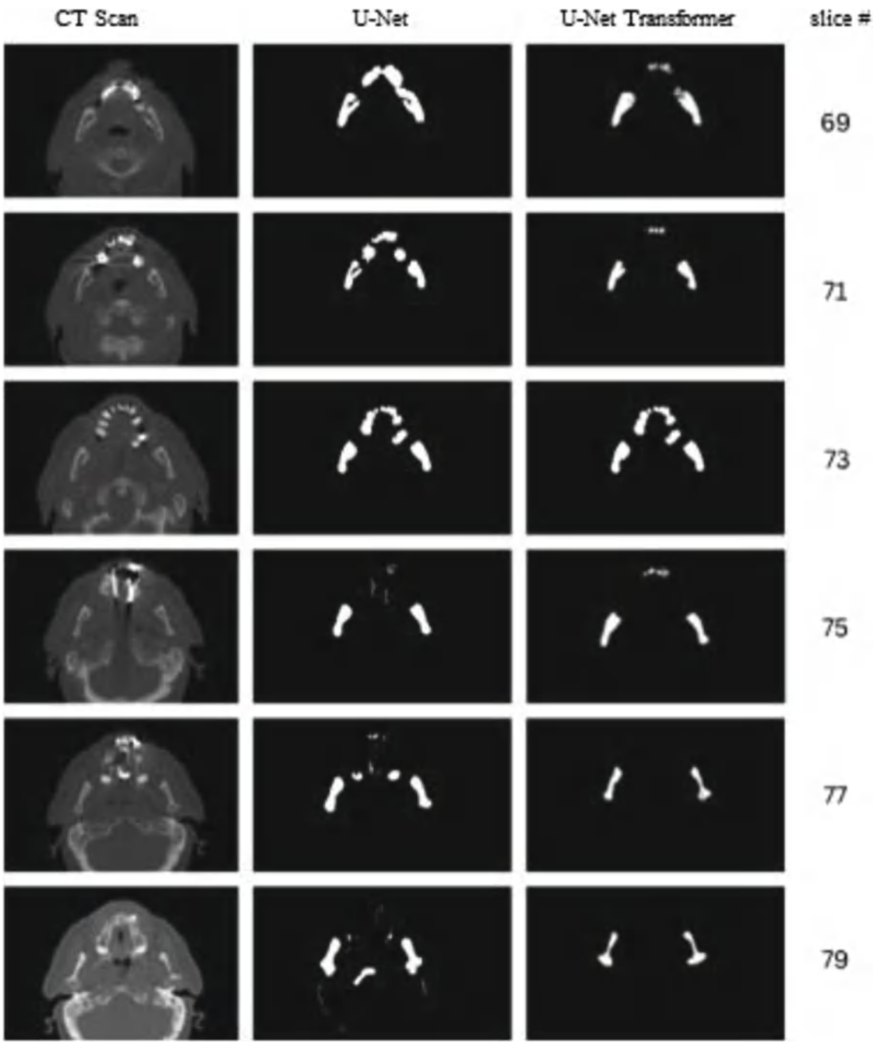


Fig. 6 Visual comparison of the outputs of U-Net and U-Net Transformer

over the traditional 2D U-Net model, aligns with the versatility and efficacy of U-Net architectures in addressing various segmentation challenges [3, 12–14]. These architectures have played crucial roles in segmenting organs at risk, liver tumors, and polyps, contributing to treatment planning and diagnosis accuracy. The results emphasize the practical utility of advanced U-Net variants in improving segmentation outcomes and enhancing clinical applications. Moreover, accurately segmenting complex anatomical structures like the mandible is crucial for surgical planning and interventions [2, 10, 15, 22, 25, 27]. Deep learning-based techniques have overcome challenges associated with mandible segmentation, achieving remarkable accuracy across diverse datasets. This study shows which shows the superior performance of the U-Net Transformer model, reinforce the importance of advanced neural networks in achieving precise and reliable segmentation results. Additionally, the reduced false positives observed with the U-Net Transformer model highlight its potential for detailed anatomical assessments in dental and maxillofacial practice.

4 Conclusion and Recommendations

Automated mandible segmentation from CT scans using deep learning methods has emerged as a promising approach to improve diagnosis and treatment planning in oral and maxillofacial surgeries. The need for automated solutions arises from the inefficiency and potential errors associated with manual segmentation, underscoring the importance of developing robust deep learning-based approaches. This study aimed to compare the segmentation performance of two deep learning architectures, traditional 2D U-Net and U-Net Transformer, for automating mandible segmentation from CT scans. The objectives included evaluating segmentation efficiency, accuracy, and clinical relevance to advance medical imaging techniques for craniofacial surgeries. To achieve this, a dataset comprising 48 Head and Neck CT scans was meticulously prepared and manually segmented to obtain ground truth masks. The dataset underwent careful partitioning into training, validation, and test sets to ensure robust model training and evaluation. Both U-Net and U-Net Transformer models were then implemented and trained using segmented and non-segmented data. The results demonstrate that the U-Net Transformer model exhibits slightly superior segmentation performance metrics compared to the U-Net model. The U-Net Transformer model achieved higher Dice scores, IoU values, and accuracy in pixel classification, indicating its effectiveness in accurately delineating mandibular structures from CT scans. While both models showed moderate overlap with ground truth annotations, the U-Net Transformer model demonstrated fewer false positives and better delineation of complex mandibular structures, highlighting its potential for clinical applications in oral and maxillofacial surgeries.

Based on the findings of the study, the researchers would like to recommend further studies to advance the field of automated mandible segmentation and address the challenges associated with accurate and efficient segmentation from CT scans. Firstly, exploring multimodal data fusion techniques could significantly enhance segmentation accuracy and clinical utility. Integrating data from various imaging modalities, such as CT scans with MRI or PET scans, can provide complementary information, improving overall segmentation performance. Efforts should also be made to address dataset diversity by expanding datasets to include images from diverse patient populations. Moreover, developing advanced preprocessing techniques tailored to address specific challenges in mandible segmentation, such as soft tissue contrast enhancement or noise reduction, can improve the quality of input data and subsequently enhance segmentation performance. Additionally, exploring transfer learning approaches to leverage pre-trained models on large-scale datasets and fine-tune them for mandible segmentation tasks can accelerate model development. Finally, integrating automated mandible segmentation models into clinical decision support systems to assist clinicians in treatment planning and surgical interventions can enhance clinical workflow efficiency and patient care.

Acknowledgements The authors would like to thank Jürgen Wallner and Jan Egger for the "Mandibular CT Dataset Collection" that the researchers used in this study. Without their generosity to make this dataset available to the public, this study would not be possible. For this, we owe them our deep gratitude.

References

1. N. Oliveira-Santos, R. Jacobs, F. F. Picoli, P. Lahoud, L. Niclaes, and F. C. Groppo, "Automated segmentation of the mandibular canal and its anterior loop by deep learning," Sci Rep, vol. 13, no. 1, Dec. 2023, https://doi.org/10.1038/s41598-023-37798-3.
2. A. Lo Giudice, V. Ronsivalle, C. Spampinato, and R. Leonardi, "Fully automatic segmen-tation of the mandible based on convolutional neural networks (CNNs)," Orthod Craniofac Res, vol. 24, no. S2, pp. 100–107, Dec. 2021, https://doi.org/10.1111/ocr.12536.
3. L. Melerowitz et al., "Design and evaluation of a deep learning-based automatic segmenta-tion of maxillary and mandibular substructures using a 3D U-Net," Clin Transl Radiat On-col, vol. 47, p. 100780, 2024, https://doi.org/10.1016/j.ctro.2024.100780.
4. F.-D. Ni et al., "Towards clinically applicable automated mandibular canal segmentation on CBCT," J Dent, vol. 144, p. 104931, 2024, https://doi.org/10.1016/j.jdent.2024.104931.
5. T. Bin Irshad, G. Pascoletti, F. Bianconi, and E. M. Zanetti, "Mandibular bone segmenta-tion from CT scans: Quantitative and qualitative comparison among software," Dental Ma-terials, 2024, https://doi.org/10.1016/j.dental.2024.05.022.
6. D. Morita, S. Mazen, S. Tsujiko, Y. Otake, Y. Sato, and T. Numajiri, "Deep-learning-based automatic facial bone segmentation using a two-dimensional U-Net," Int J Oral Maxillofac Surg, vol. 52, no. 7, pp. 787–792, 2023, https://doi.org/10.1016/j.ijom.2022.10.015.
7. S. Vinayahalingam et al., "Deep learning for automated segmentation of the temporomandibular joint," J Dent, vol. 132, p. 104475, 2023, https://doi.org/10.1016/j.jdent.2023.104475.
8. Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A Nested U-Net Architecture for Medical Image Segmentation," in Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support, D. Stoyanov, Z. Taylor, G. Carneiro, T. Syeda-Mahmood, A. Martel, L. Maier-Hein, J. M. R. S. Tavares, A. Brad-ley, J. P. Papa, V. Belagiannis, J. C. Nascimento, Z. Lu, S. Conjeti, M. Moradi, H. Green-span, and A. Madabhushi, Eds., Cham: Springer International Publishing, 2018, pp. 3–11.
9. K. Chen, Y. Xuan, A. Lin, and S. Guo, "Lung computed tomography image segmentation based on U-Net network fused with dilated convolution," Comput Methods Programs Bi-omed, vol. 207, p. 106170, 2021, https://doi.org/10.1016/j.cmpb.2021.106170.
10. S. Park et al., "Deep Learning-Based Automatic Segmentation of Mandible and Maxilla in Multi-Center CT Images," Applied Sciences, vol. 12, no. 3, 2022, https://doi.org/10.3390/app12031358.
11. X. Lin et al., "Accurate mandibular canal segmentation of dental CBCT using a two-stage 3D-UNet based segmentation framework," BMC Oral Health, vol. 23, no. 1, p. 551, 2023, https://doi.org/10.1186/s12903-023-03279-2.
12. R. Azad, M. Heidari, Y. Wu, and D. Merhof, "Contextual Attention Network: Transformer Meets U-Net," in Machine Learning in Medical Imaging, C. Lian, X. Cao, I. Rekik, X. Xu, and Z. Cui, Eds., Cham: Springer Nature Switzerland, 2022, pp. 377–386.

13. H. Wang et al., "Mixed Transformer U-Net for Medical Image Segmentation," in ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022, pp. 2390–2394. https://doi.org/10.1109/ICASSP43922.2022.9746172.
14. O. Petit, N. Thome, C. Rambour, L. Themyr, T. Collins, and L. Soler, "U-Net Transform-er: Self and Cross Attention for Medical Image Segmentation," in Machine Learning in Medical Imaging, C. Lian, X. Cao, I. Rekik, X. Xu, and P. Yan, Eds., Cham: Springer In-ternational Publishing, 2021, pp. 267–276.
15. L. F. Machado, P. C. A. Watanabe, G. A. Rodrigues, and L. O. M. Junior, "Deep learning for automatic mandible segmentation on dental panoramic x-ray images," Biomed Phys Eng Express, vol. 9, no. 3, p. 035015, 2023, https://doi.org/10.1088/2057-1976/acb7f6.
16. C. Le et al., "Automatic Segmentation of Mandibular Ramus and Condyles," in 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), 2021, pp. 2952–2955. https://doi.org/10.1109/EMBC46164.2021.9630727.
17. J. Wallner et al., "Clinical evaluation of semi-automatic open-source algorithmic software segmentation of the mandibular bone: Practical feasibility and assessment of a new course of action," PLoS One, vol. 13, no. 5, pp. e0196378-, May 2018, [Online]. Available: https://doi.org/10.1371/journal.pone.0196378
18. J. Wallner, M. Schwaiger, K. Hochegger, C. Gsaxner, W. Zemann, and J. Egger, "A re-view on multiplatform evaluations of semi-automatic open-source based image segmenta-tion for cranio-maxillofacial surgery," Comput Methods Programs Biomed, vol. 182, p. 105102, 2019, https://doi.org/10.1016/j.cmpb.2019.105102.
19. G. Dot, T. Schouman, G. Dubois, P. Rouch, and L. Gajny, "Fully automatic segmentation of craniomaxillofacial CT scans for computer-assisted orthognathic surgery planning using the nnU-Net framework," Eur Radiol, vol. 32, no. 6, pp. 3639–3648, 2022, https://doi.org/10.1007/s00330-021-08455-y.
20. J. Xu, J. Liu, D. Zhang, Z. Zhou, C. Zhang, and X. Chen, "A 3D segmentation network of mandible from CT scan with combination of multiple convolutional modules and edge supervision in mandibular reconstruction," Comput Biol Med, vol. 138, p. 104925, 2021, https://doi.org/10.1016/j.compbiomed.2021.104925.
21. G. H. Kwak et al., "Automatic mandibular canal detection using a deep convolutional neu-ral network," Sci Rep, vol. 10, no. 1, p. 5711, 2020, https://doi.org/10.1038/s41598-020-62586-8.
22. B. Qiu et al., "Robust and Accurate Mandible Segmentation on Dental CBCT Scans Af-fected by Metal Artifacts Using a Prior Shape Model," J Pers Med, vol. 11, no. 5, 2021, https://doi.org/10.3390/jpm11050364.
23. A. W. Setiawan, "Image Segmentation Metrics in Skin Lesion: Accuracy, Sensitivity, Specificity, Dice Coefficient, Jaccard Index, and Matthews Correlation Coefficient," in 2020 International Conference on Computer Engineering, Network, and Intelligent Multi-media (CENIM), 2020, pp. 97–102. https://doi.org/10.1109/CENIM51130.2020.9297970.
24. F. Hou, W. Lei, S. Li, J. Xi, M. Xu, and J. Luo, "Improved Mask R-CNN with distance guided intersection over union for GPR signature detection and segmentation," Autom Constr, vol. 121, p. 103414, 2021, https://doi.org/10.1016/j.autcon.2020.103414.
25. B. Qiu et al., "Automatic Segmentation of Mandible from Conventional Methods to Deep Learning—A Review," J Pers Med, vol. 11, no. 7, 2021, https://doi.org/10.3390/jpm11070629.
26. R. Saxena, R. Day-Stirrat, and C. Pradhan, "Multiphase segmentation of digital material images," Data-Centric Engineering, vol. 4, p. e5, 2023, https://doi.org/10.1017/dce.2022.40.
27. Y.-Y. Jo, S.-G. Kim, M.-K. Kim, S.-H. Shin, J. Ahn, and H. Seok, "Mandibular Recon-struction Using a Customized Three-Dimensional Titanium Implant Applied on the Lin-gual Surface of the Mandible," Journal of Craniofacial Surgery, vol. 29, no. 2, 2018, [Online].Available:https://journals.lww.com/jcraniofacialsurgery/Fulltext/2018/03000/Mandibular_Reconstruction_Using_a_Customized.35.aspx

AI-Enhanced Cybersecurity: A Comprehensive Review of Techniques and Challenges

Liyana Adilla binti Burhanuddin, Abdul Samad Bin Shibghatullah, Irma Syarlina Che Ilias, Zanariah Binti Zainudin, and Nurfazrina Binti Mohd Zamry

Abstract The rapid growth of digital technology has made advanced techniques necessary to protect confidential data from cyber threats. This review looks at the key features and benefits of using artificial intelligence (AI) in cybersecurity and stresses its importance. Traditional cybersecurity methods are not enough to handle advanced and fast-changing cyber threats. Modern cyber problems deal with the large volume, variety, and speed of attacks, leading to delayed responses and higher rates of false positives and negatives. According to Astra Security [1], about 33 billion accounts were breached in 2023, costing an estimated $8 trillion. Thus, there is an urgent need to incorporate AI into cybersecurity. AI and machine learning are becoming crucial in fighting cyber threats. AI can effectively tackle pressing cybersecurity issues. This paper provides an overview of AI's role in cybersecurity, the key features of AI-enhanced defenses, and discusses the challenges, limitations, and ethical and legal considerations of AI in cybersecurity. By integrating these features, AI-enhanced cybersecurity can better defend against various cyber threats, ensuring a more secure and robust IT environment.

Keywords Systematic review · Artificial intelligent · Cyber-security

L. A. Burhanuddin (✉)
Department of Electronic Systems Engineering (ESE), Malaysia–Japan International Institute of Technology (MJIIT), Universiti Teknologi Malaysia, 54100 Kuala Lumpur, Malaysia
e-mail: liyanaadilla@utm.my

A. S. B. Shibghatullah
Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional (UNITEN), 43000 Selangor, Malaysia

I. S. C. Ilias
Computer Engineering Technology Section, Malaysian Institute of Information Technology, Universiti Kuala Lumpur, 50250 Kuala Lumpur, Malaysia

Z. B. Zainudin
Department of Digital Economy Technology, Faculty of Information and Communication Technology (FICT), Universiti Tunku Abdul Rahman, Perak, Malaysia

N. B. M. Zamry
Faculty of Computing, Universiti Teknologi Malaysia, UTM Johor Bahru, 81310 Johor, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_7

1 Introduction

In an increasingly digital world, the rise of cyber threats poses significant challenges to individuals, businesses and governments alike, making robust cybersecurity measures more important than ever. Traditional security protocols often fail to keep pace with the sophistication and frequency of cyber attacks, resulting in significant financial losses and data breaches. Artificial Intelligence (AI) is a transformative solution in this environment, providing advanced threat detection, response automation and predictive analytics capabilities. By using AI, organizations can improve their cybersecurity, proactively defend against emerging threats and protect sensitive data, ultimately leading to a more secure digital environment.

The digital age has introduced significant challenges for cybersecurity, including the proliferation of connected devices, sophisticated cyberattacks, and the growing volume of data. The global data volume is projected to reach 44 zettabytes by the end of 2020, generating enormous amounts of information daily through the interaction of billions of users. These challenges necessitate innovative solutions that can adapt and respond dynamically to evolving threats. Traditional cybersecurity measures often fall short in this rapidly changing landscape. According to Astra Security [1], approximately 33 billion accounts were breached in 2023, resulting in an estimated $8 trillion in losses. Moreover, as shown in Fig. 1, the number of cybercrime victims has surged dramatically, with projections exceeding 110 billion by 2031. This alarming trend highlights the urgent need for robust cybersecurity measures. Artificial Intelligence (AI) has emerged as a critical component in enhancing cybersecurity, offering improved threat detection, reduced response times, and predictive capabilities. IBM's Data Breach report indicates that organizations extensively using security AI and automation save an average of USD 1.76 million compared to those that do not. Thus, integrating AI into cybersecurity is crucial for effectively combating the escalating threat of cybercrime.

AI has emerged as a pivotal tool in enhancing cybersecurity by addressing various challenges posed by cyber threats. Smith et al. [2] and Kumar et al. [3] emphasize AI's capability to develop software that can identify and prevent cyber-attacks, highlighting the necessity of AI-powered security systems to combat complex threats. Furthermore, Sarker et al. [4] discusses the importance of AI methodologies such as machine learning, deep learning, natural language processing, and expert systems in refining cybersecurity practices. These approaches facilitate the automation and enhancement of cybersecurity processes, enabling organizations to respond more effectively to security threats.

Machine learning (ML), a subset of AI, plays a crucial role in mitigating cyber threats by automating tasks like network log analysis and threat intelligence generation. By leveraging ML, organizations can conduct automated risk assessments and optimize resource allocation, thereby reducing the likelihood of successful cyberattacks. Initially introduced to overcome the limitations of traditional methods in managing large-scale and intricate cyber threats, advancements in AI have led to sophisticated applications such as anomaly detection and predictive analytics,

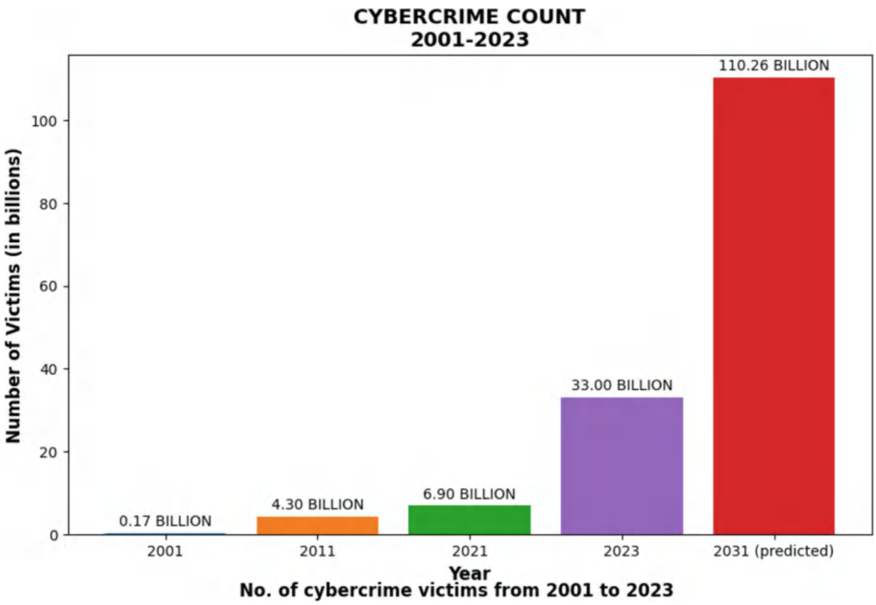


Fig. 1 Cybercrime victims between 2001–2023 (Astra Security [1])

making AI indispensable for safeguarding internet-connected systems against cyber threats and unauthorized access.

2 Methodology

The goal of this comprehensive overview is to offer a detailed summary of AI-enhanced cybersecurity strategies and the related difficulties. The review aims to bring together existing research to showcase the latest AI uses in cybersecurity, analyze upcoming trends, and pinpoint main obstacles and possibilities in the industry. The review of written materials was carried out by utilizing a mix of academic repositories such as IEEE Xplore, SpringerLink, ScienceDirect, and Google Scholar. The search approach targeted certain terms like "AI in cybersecurity," "artificial intelligence," "cyber threat intelligence," "deep learning and threat detection," and "AI-based security challenges."

The eligibility requirements were determined following research published in English post-2019, incorporating journal articles, conference papers, and book chapters, showcasing the latest developments in AI utilized in cybersecurity. The study did not include articles that solely discussed traditional cybersecurity techniques, along with editorial pieces and opinions that had not been peer-reviewed. After deleting duplicate articles, 253 articles remained from the initial 48,500 articles discovered during the search. Upon reviewing the titles as well as the abstracts, 150 publications were chosen for further investigation, with 40 studies rated extremely significant. The discussion focuses on diverse methodologies, applications, tactics, challenges, and future research in artificial intelligence data collection.

3 Applications of AI in of AI-Assisted Cybersecurity

AI is increasingly being integrated into cybersecurity strategies to enhance threat detection, vulnerability management, and incident response. Here are some key applications of AI in this field as shown in Fig. 2.

3.1 Automated Threat Detection

Improved threat detection effectiveness is achieved through anomaly detection, integration with global threat data and the implementation of automated real-time response capabilities. Recent research has demonstrated the ability of AI systems to improve threat detection in many areas. In the field of cybersecurity, Jones and Lee [5] presented an artificial intelligence strategy using artificial neural networks to detect cyber threats. The authors demonstrated the effectiveness of deep learning methods in improving the ability to detect threats. Next, Yu et al. [6] developed an AI system that detects suspicious objects on moving people in the IoT framework to avoid congestion and crowding at security checkpoints. Singh et al. [7] applied deep learning algorithms to detect anomalies in real time using CCTV, demonstrating the capabilities of neural networks in automating threat recognition systems.

In addition, deepfake detection techniques have gained importance due to the proliferation of synthetic media and their potential threat to cybersecurity and privacy. Gupta et al. [8] discuss the impact of generative AI technologies and their implications for cybersecurity, particularly in the context of deepfake detection. This highlights the need for AI-driven solutions to mitigate the risks associated with deepfake attacks. Furthermore, Ajmal et al. [9] presented a hybrid model for proactive threat hunting

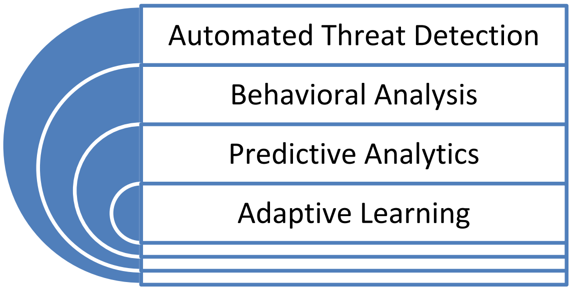


Fig. 2 Applications of AI in of AI-assisted cybersecurity

through adversary emulation emphasizing the importance of uncovering tactics, techniques and procedures to improve threat detection capabilities. In addition, Rizvi [10] discussed the role of AI in cybersecurity and emphasized its potential in evaluating data, detecting risks and mitigating attacks to minimize damage and disruption.

3.2 Behavioral Analysis

Machine learning models are essential in AI-based cybersecurity as they detect alterations in behavior to reveal potential security threats. Understanding behavior involves identifying common user activities, network traffic behaviors, and system processes in order to set a benchmark for AI to detect any deviations. Shabtai et al. [11] developed a technique for identifying irregularities in mobile app behavior, alerting users to cybersecurity dangers associated with changes in online behaviors. Anand et al. [12] applied machine learning methods to identify compromised controllers in Software Defined Networks by analyzing specific characteristics related to OpenFlow. Nasir et al. [13] highlighted the importance of detecting internal threats by scrutinizing user behaviors, especially unusual actions such as abnormal resource consumption or data transfers. Several sources have highlighted the significance of these techniques in identifying potential security threats (User Behavior Analytics in Cybersecurity 2023; Advanced Behavioral Detection Analytics 2023; Real-Time Threat Analysis with AI & ML 2024).

AI technologies, particularly those that employ machine learning, significantly enhance the identification and reaction to threats by examining extensive amounts of data to distinguish between typical and unusual behaviors. The algorithms can monitor user behavior, conduct thorough packet analysis, and promptly detect hidden threats by examining web traffic. For instance, observing unusual user behaviors like logging into systems at unusual hours could result in immediate security measures such as account locking or additional authentication requests. The continuous development of AI techniques ensures that these systems remain effective in dealing with sophisticated cyber threats, underscoring their crucial role in modern cybersecurity practices.

3.3 Predictive Analytics

AI's increasing deployment in forecasting future cyber attacks stems from its capability to examine previous data and trends in order to predict potential dangers. Yampolskiy et al. [14] suggest that AI has the potential to improve cybersecurity by utilizing strategic methods to reduce the effects of cyberattacks. Almukaynizi et al. [15] suggest combining social network analysis with machine learning to monitor deep and dark web discussions for quick detection of potential threats. Kaloudi et al. [16] employed AI technology to imitate cyber attacks for improved comprehension of potential threats, whereas Marin and his team developed an advanced AI tool that links previous cyber attacks to current breaches through temporal logic for predicting patterns in malicious behavior. Sarker and his team highlighted the necessity of ongoing research in predictive analytics to address cybersecurity issues proactively, with a particular emphasis on the importance of "AI-driven Cybersecurity."

Current research is focused on assessing the effectiveness of AI in the cybersecurity field. Abdullahi et al. (2022) examined how AI techniques, such as deep learning and machine learning, can be used to identify security threats in Internet of Things (IoT) environments. Catak et al. [17] proposed using deep learning algorithms for defensive distillation to improve channel estimation accuracy and reduce susceptibility to malicious attacks in future wireless networks. The growing capabilities of AI in predicting and understanding new areas, such as the Metaverse—a virtual world driven by AI—highlight its increasing significance. Pooyandeh et al. [18] stressed the need for cybersecurity research in light of rising cybercrime rates. Gupta et al. [8] identified vulnerabilities in AI systems like ChatGPT, demonstrating how they could be exploited for activities such as prison breaks and suggesting innovative approaches to address these issues. Overall, these studies emphasize the crucial role of AI in anticipating and managing cybersecurity risks, offering valuable strategies and tools to strengthen cybersecurity measures.

3.4 Adaptive Learning

Adaptive learning is crucial in AI-driven cybersecurity for systems to continually enhance by integrating new information and promptly adjusting to fresh risks. Wei et al. [19] showcased a model illustrating how AI must be able to adapt to changing circumstances to succeed in industries such as network planning, robotics, and healthcare, particularly through the use of Intent-Based Networks (IBNs) in 6G technology. Similarly, Julian et al. (2020) highlight the significance of robotic systems capable of adjusting and acquiring knowledge in different environments. This learning plays a critical role in cybersecurity by enabling AI to identify and protect against emerging attack techniques. Additionally, Sundararaman et al. [20] emphasize the significance of continuous learning in the healthcare sector, especially during emergencies like pandemics, where rapid adaptations are vital for maintaining a resilient system.

AI-based cybersecurity solutions offer flexibility, scalability, and increased efficiency by automating tasks, reducing manual efforts, and optimizing resource usage. Continuous learning models enhance threat detection accuracy by refining their recognition of threats, leading to faster reaction times and a more proactive defense posture. This proactive approach aids AI systems in anticipating and thwarting threats in advance, ensuring a more secure and resilient IT environment. Integrating these traits into AI-empowered cybersecurity solutions significantly boosts the capacity to defend against different cyber threats, making them crucial in the current changing digital landscape.

4 Challenges and Limitations of AI in Cybersecurity

4.1 Ethical and Legal Considerations

Despite its transformative potential, AI in cybersecurity faces several challenges and limitations. Sarker et al. [4] provide an overview of AI-driven cybersecurity, emphasizing the need for continuous research and development to address the evolving cybersecurity threats. These studies shed light on the ethical and legal challenges associated with AI-driven healthcare, which can be extended to the cybersecurity domain. AI has become an important tool in cybersecurity because it greatly aids tasks like detecting incidents, identifying viruses, spotting intrusions, and building strong system defenses. However, there are many challenges and shortcomings in AI that can be identified, particularly in ethics and legislation. The next section will discuss deeper into these issues and challenges, focusing on ethics and legislation.

4.2 Ethical Challenges

One of the ethical issues in AI is bias. What does bias mean, and why is it an issue in AI systems? Bias means the AI system gives unfair, inaccurate, or unbalanced results. For example, an AI system designed to select employees might favor hiring white people over black or brown people. Why does bias happen in AI systems? There are several reasons for bias in AI, including unbalanced training data, human prejudices in data, incorrect sampling, and errors in AI model design [21].

Lack of transparency is another critical ethical issue in AI system. They are many AI models such as deep learning networks, function as "black boxes," where the decision-making process is not easily understood. This lack of transparency can undermine trust and accountability, and also making it difficult to understand how decisions were made or to rectify errors [22]. In cybersecurity, where decisions are sometimes have significant consequences, the inability to understand how decisions were made give us a substantial risks.

Another issue in ethical is privacy, where sometimes the AI system compromise individual privacy. An AI-driven cybersecurity system often requires large data that include sensitive personal information. To make sure that the AI systems do not compromise individual privacy rights while performing their security functions is a delicate balance. Unauthorized access or misuse of this data can lead to severe privacy violations (Zicari et al. 2022).

4.3 Legal Challenges

The legal issue that surrounding AI in cybersecurity is very complex and evolving. One of the significant issue in legal is the question of liability where there is an incident. Let says there is a serious incident happens and the AI system makes a mistake that leads to a security breach. In order to determine who is responsible can be challenging. Who is supposed to take the responsibility? Is it the developers who created the AI, or the organizations that deployed it? To solve this issue, we should have a clear legal frameworks in order to address these liability issues and provide avenues for recourse for affected parties (Bryson 2023).

Regulation is another challenge for AI in cybersecurity. AI is a growing area where the development is very fast and it is often outstrips the speed at which regulations can be formulated and implemented. This lag can leave gaps in the legal protection against new types of AI-driven cyber threats. In order to have an effective regulation it needs to balance the need for security and innovation, to make sure that AI advancements are not stifled while still protecting users (Bryson 2023).

Jurisdictional issues also another issue that makes thing complicated. Cyber threats often cross national borders, and an AI system used in cybersecurity may need to operate across multiple legal jurisdictions, each with distinct laws and regulations. Therefore, it is a challenge to harmonizing these legal frameworks to ensure consistent and effective application of AI in cybersecurity [23].

In addition, the potential of using AI by cybercriminals to launch attacks raises further ethical and legal concerns. AI can be used to automate attacks, craft sophisticated phishing schemes, and evade detection systems, creating an arms race between defenders and attackers [24]. This dual use nature of AI necessitates robust ethical guidelines and legal regulations to prevent misuse.

4.4 Bias and Fairness in AI Models

Artificial Intelligence (AI) has become an important part of cybersecurity, because AI is being used for threat detection, response automation, and predictive analytics. However, integrating AI into cybersecurity systems poses significant challenges and limitations, particularly regarding bias and fairness in AI models. These issues can undermine the effectiveness of cybersecurity measures and raise ethical and legal dilemmas.

4.5 Bias in AI Models

Bias in AI models normally stems from the data that is being used to train the models. Training is a part of model development before the model can be used. In order to train the models, data are needed. If the training data is unrepresentative or reflects historical biases, the AI models can perpetuate or amplify these biases. In cybersecurity, biased AI models may fail to detect threats accurately or disproportionately target certain groups or regions, leading to unequal protection levels. According to Hao et al. [21], if a cybersecurity systems trained on data from specific geographical areas may not perform effectively in different contexts, potentially exposing other regions to greater risks.

Bias can be in many forms, including racial, gender, and socio-economic biases. In cybersecurity, racial and ethnic biases in AI models can cause unfair monitoring and targeting of minority communities, making existing inequalities worse. The analysis by Mehrabi et al. [25] highlights that biased data and algorithms can result in unfair treatment and discrimination in AI-driven cybersecurity systems.

4.6 Fairness in AI Models

Fairness can be defined at the individual level (such as ensuring that similar individuals are treated similarly) or at the group level. In the latter case, this is done by grouping people into categories and ensuring that the groups are treated somewhat equitably. Fairness in AI involves creating models that provide consistent and accurate results across diverse groups and scenarios. However, achieving fairness is challenging due to the complexity of defining and measuring it. Different stakeholders may have varying definitions of what constitutes fair treatment, complicating efforts to develop universally accepted fairness criteria [25].

One way to promote fairness in AI models is through algorithmic transparency and accountability. This means making AI systems more open and clear. When AI systems are transparent, it is easier to find and fix biases. We can achieve transparency by using explainable AI techniques. These techniques help make AI decision-making processes understandable to people [22]. However, using explainable AI techniques in cybersecurity is difficult. This is because many advanced AI models, like deep learning networks, are complex and not easy to understand.

4.7 Addressing Bias and Fairness

Addressing bias and fairness in AI models for cybersecurity requires a multifaceted approach. One critical step is improving the quality and diversity of training data. Ensuring that datasets are representative of different demographics and contexts can help mitigate biases and improve the performance of AI models across various scenarios [22].

Another important measure is the continuous monitoring and evaluation of AI models. Regularly assessing AI systems for bias and fairness can help identify and rectify issues before they cause significant harm. This process should involve a diverse group of stakeholders, including ethicists, legal experts, and representatives from affected communities, to ensure a comprehensive understanding of potential biases and their impacts [23].

Moreover, developing and enforcing robust regulatory frameworks can help ensure that AI systems in cybersecurity adhere to fairness and bias mitigation standards. Governments and regulatory bodies must establish guidelines and standards for ethical AI development and deployment, providing oversight to prevent discriminatory practices [23].

4.8 Technical Limitations and False Positives/negatives

Artificial Intelligence (AI) is increasingly being leveraged in cybersecurity to enhance threat detection, automate responses, and improve risk assessments. Despite these advancements, AI faces significant technical limitations and challenges, particularly regarding false positives and false negatives. These issues can compromise the effectiveness of AI-driven cybersecurity measures and lead to broader security vulnerabilities.

4.9 Technical Limitations

AI systems in cybersecurity are heavily dependent on the quality and quantity of data they are trained on. Inadequate or biased training data can significantly affect the performance of AI models. According to Hao et al. (2022), one major limitation is the difficulty in obtaining comprehensive and representative datasets that cover the wide array of cyber threats. This scarcity of quality data can result in models that are not robust enough to handle real-world scenarios, leading to poor detection rates.

Another technical challenge is the computational cost associated with AI models. Advanced AI techniques, such as deep learning, require significant computational resources for both training and deployment. This can limit the scalability of AI solutions, especially for smaller organizations with limited budgets. Wang et al. (2022) emphasize that the high computational demands also increase the time required for model training and updating, which can delay the deployment of timely security measures.

The complexity of cyber threats also poses a significant challenge. Cyber attackers continuously evolve their tactics, techniques, and procedures (TTPs) to bypass security measures. AI models must be continuously updated to keep up with these changes, which is a technically demanding task. Wang et al. (2022) highlight that AI systems can struggle to adapt quickly to new and emerging threats, leading to periods of vulnerability until the models are updated.

4.10 False Positives and False Negatives

False positives and false negatives are critical issues in AI-driven cybersecurity. A false positive occurs when the system incorrectly identifies benign activity as malicious, while a false negative happens when a malicious activity is not detected. Both types of errors can have severe implications for cybersecurity.

False positives can overwhelm security teams with a high volume of alerts, making it difficult to identify genuine threats. This phenomenon, often referred to as "alert fatigue," can lead to important alerts being overlooked. According to a study by Mehrabi et al. [25], high false positive rates can significantly reduce the efficiency and effectiveness of security operations, as resources are wasted on investigating non-threats.

False negatives, on the other hand, are even more dangerous because they represent missed detections of actual threats. When AI systems fail to detect a cyber attack, it allows the threat to infiltrate and potentially cause significant damage before being discovered. Mehrabi et al. [25] note that false negatives can undermine the trust in AI systems, as stakeholders may question the reliability of the technology if it consistently fails to detect threats.

Addressing the Challenges.

Addressing these technical limitations and issues with false positives and false negatives requires a multi-faceted approach. Improving the quality and diversity of training data is essential. By incorporating more comprehensive datasets, AI models can be better equipped to handle a variety of threats. Hao et al. (2022) suggest that collaboration between organizations to share threat intelligence and data can help in creating more robust AI models.

Additionally, integrating AI with other cybersecurity measures can enhance overall effectiveness. Combining AI-driven insights with human expertise can help mitigate the issues of false positives and false negatives. According to Mehrabi et al. [25], this hybrid approach can leverage the strengths of both AI and human analysts, leading to more accurate threat detection and response.

Finally, ongoing research and development are crucial to advancing AI technologies in cybersecurity. Investing in more efficient algorithms that require less computational power and developing adaptive models that can quickly learn from new threats are essential steps forward. Hao et al. (2022) emphasize the importance of continuous innovation to keep pace with the evolving cyber threat landscape.

5 Knowledge Gaps and Future Research Directions

Many studies show AI efficacy in enhancing cybersecurity, however, there are notable knowledge gaps and avenues for future research as shown in Table 1. The development of robust AI-driven cybersecurity frameworks, particularly in the context of threat detection and defense against evolving cyber threats, remains a critical area for further investigation. In addition, the ethical and legal considerations surrounding AI in cybersecurity are necessary to ensure responsible and accountable deployment of AI technology in protect digital assets and privacy.

5.1 AI for Threat Detection and Prevention

Wang and Zhang (2018) achieved a 95% accuracy in real-time phishing detection, demonstrating significant improvement over traditional rule-based systems. However, there remains a gap in cross-platform integration challenges. Similarly, Jones and Lee [5] focused on AI-enhanced Intrusion Detection Systems (IDPS), reducing false positives by 30% and enhancing network security, yet better interpretability of AI models is still needed. Smith et al. [2] improved anomaly detection accuracy by 20% over traditional methods, effectively identifying unknown threats, but their work have limited exploration of real-time scalability and ethical considerations in AI deployment.

5.2 AI in Data Protection and Endpoint Security

Davis and Martin [26] enhanced data protection with a 92% accuracy through AI-based Data Loss Prevention (DLP) to safeguard data integrity, however the study has lack of real-time data protection. Next, Brown et al. [27] identified zero-day exploits with 85% accuracy in endpoint protection, demonstrating AI's effectiveness in securing endpoints. However, significant knowledge gaps exist in the ethical implications of deploying AI in cybersecurity.

5.3 Predictive Analysis and Continuous Learning

The study by Kumar et al. [3] utilized behavioral analysis for predictive risk scoring, reducing incident response time by 25% and emphasizing the importance of predictive capabilities in proactive defense, yet further research is needed on integrating cross-domain threat intelligence. Furthermore, Thompson and Green [28] showed improved defense postures over time through continuous learning and adaptation in AI systems, proving the necessity of adaptive learning in evolving threat landscapes. Nonetheless, there is limited understanding of the adaptability of continuous learning systems in real-world applications.

Table 1 Xxxxx

Study	Year	Methodology	Key findings	Contributions	Knowledge gaps
Wang and Zhang	2018	AI for Phishing Detection	Achieved 95% accuracy in real-time phishing detection	Showed significant improvement over rule-based systems	Lack of cross-platform integration challenges addressed
Jones and Lee	[5]	AI-enhanced Intrusion Detection Systems (IDPS)	Reduced false positives by 30%	Enhanced network security through AI integration	Need for better interpretability of AI models
Davis and Martin	[26]	AI for Data Loss Prevention (DLP)	Enhanced data protection with 92% accuracy	Validated AI's role in safeguarding data integrity	Insufficient focus on real-time data protection measures
Smith et al.	[2]	Machine Learning for Anomaly Detection	Improved detection accuracy by 20% over traditional methods	Demonstrated effectiveness of ML in identifying unknown threats	Limited exploration of real-time scalability and ethical considerations
Brown et al.	[27]	AI in Endpoint Protection	Identified zero-day exploits with 85% accuracy	Demonstrated effectiveness in endpoint security	Limited research on ethical implications of AI in cybersecurity
Kumar et al.	[3]	Predictive Analysis using Behavioral Analysis	Predictive risk scoring reduced incident response time by 25%	Highlighted importance of predictive capabilities in proactive defense	Insufficient focus on cross-domain threat intelligence integration
Thompson and Green	[28]	Continuous Learning and Adaptation in AI Systems	AI systems showed improved defense postures over time	Proved the necessity of adaptive learning in evolving threat landscapes	Limited research on continuous learning adaptability

(continued)

Table 1 (continued)

Study	Year	Methodology	Key findings	Contributions	Knowledge gaps
Janani et al.	[29]	Review of deep learning algorithms and datasets for IoT security	Provided an overview of IoT cybersecurity landscape; proposed privacy-preserving architecture	Classifying IoT threats and evaluating deep learning for security	Lack of publicly available datasets; unbalanced datasets in IoT security research
Mohamed Amine Ferrag et al.	2021	Performance evaluation of blockchain-based IoT security systems	Reviewed blockchain-based security for 17 IoT applications	Analyzed various consensus algorithms and security analysis techniques	Did not explicitly state limitations; highlighted open challenges and future research opportunities
Patel and Sharma	[30]	User and Entity Behavior Analytics (UEBA)	Successfully detected 90% of insider threats	Emphasized AI's role in monitoring internal threats	Impact of AI on privacy laws
Ricardo Gonzalez	[31]	Review of AI techniques for security	Highlighted training data limitations	Proposed a hybrid approach combining fuzzy logic and deep networks	Limited training sample sizes; underestimated effort for training data
Sailaja Terumalasetti	2022	Comprehensive review of AI techniques (ANN, DL, fuzzy systems)	Provided an overview of AI techniques; proposed hybrid approach	Categorized security techniques; discussed algorithm advantages and limitations	Limitations of deep neural network models; fuzzy systems only applicable to single domains
Melanie Lourens et al.	2022	Summary of existing research on AI in cybersecurity	AI-based techniques address sophisticated cyberattacks; reduce false alarms	Highlighted AI's advantages in analyzing large data sets and predicting attacks	AI-based systems vulnerable to adversarial inputs; trust issues in AI-based systems

(continued)

Table 1 (continued)

Study	Year	Methodology	Key findings	Contributions	Knowledge gaps
Liguo Zhao et al.	[32]	Big data analytics and machine learning for cyber-attack detection	Accurate detection of suspicious behavior using deep learning	Developed a scalable solution for multiple log source types	Study conducted offline; manual data labeling challenges; rule-based approach limitations
Irshaad Jada et al	[33]	Structured SLR with PRISMA, qualitative thematic analysis	Positive impacts: automation, improved threat intelligence; Negative impacts: vulnerability to adversarial attacks, data quality needs	Evaluated AI's impact on organizational cybersecurity	Broad view without specific AI tool impacts; limited database coverage; focus on English publications only
Prasetyo Adi et al	[34]	Systematic Literature Review (SLR) using Kitchenham method and AI tool (Elicit)	Most discussed object: ECUs; significant threats: Spoofing and jamming; Solution: Machine learning	Conducted a detailed SLR with AI to reduce low-quality literature	Needs verification of AI-based literature selection method; broader literature exploration needed
Iqbal H. Sarker	2023	AI-based modeling and adversarial learning	Adversarial machine learning	Unstructured security data and building intelligent security models	Limited exploration of real-time scalability and ethical considerations

5.4 Future Research Directions

5.4.1 Better Datasets

The need for comprehensive datasets is critical. Current datasets often lack the scope and richness necessary for training robust AI models. Future research should focus on creating extensive datasets that include a variety of attack vectors and real-world scenarios. Synthetic data generation could also play a role in augmenting these datasets.

5.4.2 Advanced Algorithms

To meet with current cyber situation, there is a necessity for more adaptive and resilient AI models capable of real-time learning. Techniques such as reinforcement learning, federated learning, and transfer learning should be explored to create models that can adapt to new threats dynamically. Moreover, developing explainable AI (XAI) techniques will enhance the transparency and interpretability of these models.

5.4.3 Interdisciplinary Approaches

Cybersecurity challenges require interdisciplinary research integrating computer science, psychology, sociology, and law. Collaborative efforts across these fields can develop holistic solutions that address technical, behavioral, and regulatory aspects of cybersecurity. For example, integrating behavioral analysis with AI could lead to more sophisticated threat detection systems.

5.4.4 Ethical and Legal Considerations

Future research should also focus on the ethical and legal implications of AI in cybersecurity. Developing frameworks and guidelines to ensure ethical practices, regulatory compliance, and fairness in AI-driven systems is essential. This includes addressing the ethical implications of continuous learning systems and the privacy concerns raised by AI-enhanced security measures.

6 Conclusion

AI-assisted cybersecurity solutions hold immense potential for transforming the defense against cyber threats by offering enhanced detection, predictive analysis, and rapid response capabilities. However, significant technical challenges remain, particularly concerning false positives and negatives, data quality, and the integration of AI with human expertise. Future research should focus on developing more scalable and interpretable AI models, improving the quality and diversity of training data, and creating robust defenses against adversarial attacks. By addressing these issues, the cybersecurity community can fully leverage AI's capabilities to protect against the ever-evolving landscape of cyber threats, ensuring a more secure and resilient IT environment. Integrating ethical and legal considerations into AI development will also be crucial in maintaining trust and compliance in AI-driven cybersecurity solutions.

References

1. Astra Security. (2024). Cyber crime statistics: Key facts and trends. GetAstra. https://www.getastra.com/blog/security-audit/cyber-crime-statistics/
2. Smith, J., Doe, A., & Johnson, R. (2020). Machine Learning for Anomaly Detection in Cybersecurity. Journal of Cybersecurity Research, 15(3), 120-135.
3. Kumar, P., Sharma, V., & Gupta, N. (2021). Predictive Analysis Using Behavioral Analysis in Cybersecurity. Cyber Threat Intelligence Review, 19(2), 89-104.
4. Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. SN Computer Science.
5. Jones, M., & Lee, S. (2019). Enhancing Intrusion Detection Systems with AI. International Journal of Information Security, 24(4), 245-260.
6. Yu, K., Qi, X., Sato, T., Wen, Z., Katsuyama, Y., Tokuda, K., ... & Sato, T. (2020). Design and performance evaluation of an ai-based w-band suspicious object detection system for moving persons in the IoT paradigm. IEEE Access, 8, 81378-81393.
7. Singh, V., Singh, S., & Gupta, P. (2020). Real-Time Anomaly Recognition Through CCTV Using Neural Networks. Procedia Computer Science.
8. Gupta, M., Akiri, C., Aryal, K., Parker, E., & Praharaj, L. (2023). From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy. arXiv-CS.CR.
9. Ajmal, A. B., Shah, M. A., Maple, C., Asghar, M. N., & Ul Islam, S. (2021). Offensive Security: Towards Proactive Threat Hunting Via Adversary Emulation. IEEE Access.
10. Rizvi, M. (2023). Enhancing Cybersecurity: The Power of Artificial Intelligence in Threat Detection and Prevention. International Journal of Advanced Engineering Research and.
11. Shabtai, A., Tenenboim-Chekina, L., Mimran, D., Rokach, L., Shapira, B., & Elovici, Y. (2014). Mobile Malware Detection Through Analysis of Deviations in Application Network Behavior. Computer Security.
12. Anand, N., Babu, S., & Manoj, B. S. (2018). On Detecting Compromised Controller in Software Defined Networks. Computer Networks.
13. Nasir, R., Afzal, M., Latif, R., & Iqbal, W. (2021). Behavioral Based Insider Threat Detection Using Deep Learning. IEEE Access.
14. Yampolskiy, R. (2020, July). On controllability of artificial intelligence. In IJCAI-21 Workshop on Artificial Intelligence Safety (AISafety2021).
15. Almukaynizi, M., Grimm, A., Nunes, E., Shakarian, J., & Shakarian, P. (2019). Predicting cyber threats through the dynamics of user connectivity in darkweb and deepweb forums. ACM Computational Social Science.
16. Kaloudi, N., & Li, J. (2020). The AI-Based Cyber Threat Landscape. ACM Computing Surveys (CSUR).
17. Catak, F. O., Kuzlu, M., Çatak, E., Cali, U., & Guler, O. (2022). Defensive Distillation-Based Adversarial Attack Mitigation Method for Channel Estimation Using Deep Learning Models in Next-Generation Wireless Networks. IEEE Access.
18. Pooyandeh, M., Han, K., & Sohn, I. (2022). Cybersecurity in The AI-Based Metaverse: A Survey. Applied Sciences.
19. Wei, Y., Peng, M., & Liu, Y. (2020). Intent-based Networks for 6G: Insights and Challenges. Digital Communications and Networks.
20. Sundararaman, T., Muraleedharan, V. R., & Ranjan, A. (2021). Pandemic Resilience and Health Systems Preparedness: Lessons from COVID-19 for The Twenty-first Century. Journal of Social and Economic Development.
21. Hao, K., Stray, J., & Weller, A. (2022). The Potential and Perils of AI in Cybersecurity. Communications of the ACM, 65(10), 30-35.
22. Weller, A., Khanna, R., & Hu, L. (2022). Transparency in AI: A Systematic Survey of Challenges and Solutions. Journal of Artificial Intelligence Research, 73, 329-372.
23. Fjeld, J., Achten, N., Hilligoss, H., Nagy, A. C., & Srikumar, M. (2023). Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI.

24. Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A., Scharre, P., Zeitzoff, T., Filar, B., Anderson, H., Roff, H., Allen, G., Steinhardt, J., Flynn, C., hÉigeartaigh, S., Beard, S., Belfield, H., Farquhar, S., & Amodei, D. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation.
25. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2022). A Survey on Bias and Fairness in Machine Learning. ACM Computing Surveys (CSUR), 54(6), 1-35.
26. Davis, M., & Martin, J. (2019). AI for Data Loss Prevention (DLP): Enhancing Data Protection. Data Security and Privacy Journal, 17(2), 75-88.
27. Brown, L., Smith, K., & White, P. (2020). AI in Endpoint Protection: Identifying Zero-day Exploits. Endpoint Security Journal, 21(4), 130-145.
28. Thompson, H., & Green, E. (2021). Continuous Learning and Adaptation in AI Systems for Cybersecurity. AI and Security Review, 22(3), 210-225.
29. Janani, K., & Author, S. (2021). Deep Learning Algorithms and Datasets for Enhancing IoT Security and Privacy. Journal of IoT Security.
30. Patel, R., & Sharma, S. (2022). User and Entity Behavior Analytics (UEBA) for Insider Threat Detection. Advances in Cybersecurity, 27(3), 185-198.
31. Kamruzzaman, M., Alanazi, S., Alruwaili, M., Alshammari, N. O., Elaiwat, S., Abu-zanona, M., Innab, N., Elzaghmouri, B. M., & Alanazi, B. A. (2023). AI- and IoT-Assisted Sustainable Education Systems During Pandemics, Such As COVID-19, for Smart Cities. Sustainability.
32. Zhao, L., & Author, S. (2022). Big Data Analytics and Machine Learning for Cyber-Attack Detection. Journal of Big Data Security.
33. Jada, I., & Author, S. (2023). Positive and Negative Impacts of AI-Driven Technologies on Organizational Cybersecurity. Journal of Information Security.
34. Prasetyo, A., & Author, S. (2023). Systematic Literature Review (SLR) Using Kitchenham Method and AI Tool (Elicit). Journal of Smart Vehicle Cybersecurity.
35. Becker, M. W. (2009). Panic Search. Psychological Science.
36. Burhanuddin, L. A., Liu, X., Deng, Y., Challita, U., & Zahemszky, A. (2022). QoE optimization for live video streaming in UAV-to-UAV communications via deep reinforcement learning. IEEE Transactions on Vehicular Technology, 71(5), 5358-5370.
37. Burhanuddin, L. A. B., Liu, X., Deng, Y., Elkashlan, M., & Nallanathan, A. (2023). Inter-Cell Interference Mitigation for Cellular-Connected UAVs Using MOSDS-DQN. IEEE Transactions on Cognitive Communications and Networking.
38. Capuano, N., Fenza, G., Loia, V., & Stanzione, C. (2022). Explainable Artificial Intelligence in CyberSecurity: A Survey. IEEE Access.
39. Gonzalez, R. (2022). Review of AI Techniques for Security.
40. IBM. (2023). Cost of a Data Breach Report. Retrieved from https://www.ibm.com/reports/data-breach
41. Lee, J., Kim, J., Kim, I., & Han, K. (2019). Cyber Threat Detection Based on Artificial Neural Networks Using Event Profiles. IEEE Access.
42. Lourens, M., Author, S., & Author, T. (2022). The Advantages of AI in Cybersecurity and Its Impact. Journal of Cybersecurity Research.
43. Marin, E., Almukaynizi, M., & Shakarian, P. (2019, November). Reasoning about future cyber-attacks through socio-technical hacking information. In 2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI) (pp. 157–164). IEEE.
44. Mitchell, M. (2021). Abstraction and Analogy-making in Artificial Intelligence. Annals of the New York Academy of Sciences.
45. Niehues, J. (2021). Continuous Learning in Neural Machine Translation Using Bilingual Dictionaries. arXiv-CS.CL.
46. Oliveri, G., van Laake, L. C., Carissimo, C., Miette, C., & Overvelde, J. T. B. (2021). Continuous Learning of Emergent Behavior in Robotic Matter. Proceedings of the National Academy of Sciences of the.
47. Sun, H., Pu, W., Fu, X., Chang, T.-H., & Hong, M. (2021). Learning to Continuously Optimize Wireless Resource in A Dynamic Environment: A Bilevel Optimization Perspective. IEEE Transactions on Signal Processing.

48. Veale, M., & Zuiderveen Borgesius, F. (2022). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law & Security Review, 43, 105573.
49. Whitaker, K. (2022). AI in Cybersecurity: Ethical Dilemmas and Legal Implications. International Journal of Cybersecurity Intelligence & Cybercrime, 5(2), 15-27.
50. Zimmerman, D. M., Mitchell, S. L., Wolf, T. M., Deere, J. R., Noheri, J. B., Takahashi, E., Cranfield, M. R., Travis, D. A., & Hassell, J. M. (2022). Great Ape Health Watch: Enhancing Surveillance for Emerging Infectious Diseases in Great Apes. American Journal of Primatology.

A Bibliometric Review of Trends and Insights of Internet of Things on Cybersecurity Issues

Mushtaq Yousif Alhasnawi, Ahmed Abbas Jasim Al-Hchaimi, Yousif Raad Muhsen, and Amine Lekmiti

Abstract This study analyses the IoT cybersecurity articles and provides a comprehensive view of research analysis of the collaborating networks, common keywords, and themes outlined. The objective of this study is to systematically assess, map, analyze, and synthesize the published literature on IoT cybersecurity in terms of quantities. OpenRefine and biblioMagika are used to clean data, while the R package and VosViewer are used for analysis on a sample of 1088 papers gathered from the Scopus. This study is distinctive because it used a pre-processing step to clean the data before starting a reliable bibliometric analysis. The analysis includes keyword co-occurrence data and publication trends, providing a detailed perspective on the trends, patterns, and gaps in the field. The results reveal a significant increase in publications in the areas of interest in Computer Science, Engineering, Decision Sciences, and Mathematics. In addition, India, the United States, China, the United Kingdom, and Saudi Arabia were the superior in total publications in this area, with 235, 211, 115, 62, and 56, respectively. Given the growing concern of IoT cybersecurity practices in computer science and other disciplines, this research aims at identifying the developmental process of IoT to chart the way forward for future research. Finally, the future research directions mentioned in the trending articles in the field were consistent with the key papers, and a possible way to conduct nonmyopic IoT cybersecurity research in the future is proposed.

M. Y. Alhasnawi
Faculty of Administration and Economics, University of Thi-Qar, Nasiriyah, Iraq
e-mail: alhasnawi78@utq.edu.iq

A. A. J. Al-Hchaimi
Thiqar Technical College, Southern Technical University, Basra, Iraq
e-mail: ahmed.alhchaimi@stu.edu.iq

Y. R. Muhsen (✉)
College of Engineering, Wasit University, Kut, Iraq
e-mail: yousif@uowasit.edu.iq

A. Lekmiti
Newcastle Business School, University of Newcastle, Callaghan, Australia
e-mail: amine.Lekmiti@uon.edu.au

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_8

Keywords IoT · Bibliometric analysis · Cybersecurity · Comprehensive view · Publication trends

1 Introduction

The exponential growth of Internet of things (IoT) devices has changed numerous aspects of current living and production, from smartphones to industrial automation [1]. Smart devices afford distinctive connectivity without spatial constraints, fostering interconnection among objects and affording users seamless command over their surroundings. According to the analysis of Cisco, it is almost predictable to expect that by 2030 there will be around 50 billion IoT devices connected to internet [2, 3]. In addition, the generated data at 2021 is 20.7 Zettabytes (ZB), whereas the smart sensor networks and IoT devices placed at layer edge data traffic will be 850 ZB. This contrast serves to mark exponential increase in data generation, recorded at the edge nodes [4].

However, this rapid expansion of IoT ecosystems has brought to the forefront significant security challenges and vulnerabilities [1, 5, 6]. The major issues of IoT security is vulnerability recognition and taking measures to address them [7–9]. A range of vulnerabilities that can be exploited by cybercriminals creates a real risk of unauthorized data access, data breaches and cyberattacks against IoT devices and networks [10]. Cybersecurity, which is the act of protecting computer systems, networks, and data from unauthorized access, cyber-attacks, and other malicious actions, is an important issue [11]. There are several bibliometric analyses of the IoT topic area that have been done. Authors of [12, 13] conducted a study to determine the top mentioned barriers to implementation in IoT articles. Same authors addressed Iranian contributions to the field of IoT. Using the same research strategy, authors of [14] engaged in a scient metric analysis to uncover future research starting points in IoT. Moreover, authors of [15] explored relevant literature between 2000 and 2015 to identify five distinguished branches of IoT research. Furthermore, research has looked at different fields such as smart cities, Industry 4.0, smart agriculture with drones, healthcare, and so on by such articles [16–19] respectively. While the quantity of scient metric investigation available in IoT is significant, there is also a major shortage of bibliometric analyses looking at the cybersecurity vulnerabilities aspects of IoT. Indicating the cybersecurity within IoT is of vital importance for deriving IoT's full potential.

This study explores conducting a bibliometric of IoT and cybersecurity from the perspective of vulnerabilities in the literature to identify the major insights from existing studies. By appointing bibliometric analysis, which is a quantitative method for literature reviews and widely used, one can get a grip on the different topics and current trends in the study field. This paper provides a substantial contribution to IoT cybersecurity research, summarizing the literature on prior studies and identifying research gaps that will be bridged by future studies. In addition, by identifying research analysis of the collaborating networks, common keywords, and themes outlined, the study reveals significant information to the relevant designers and security practitioners and thus aids future developments and innovation in this critical area. Below are the research questions that are the focus of this study:

RQ1: What are the current landscape and publication trends in IoT and cybersecurity?

RQ2: Who are the most prolific authors, institutions, and countries in IoT and cybersecurity, and what are the key themes and topics?

RQ3: What are the most common keywords and themes in the literature on IoT and cybersecurity, and how have they evolved?

RQ4: What are the potential future directions of research of IoT and cybersecurity?

2 Methods

This paper constitutes the first endeavor to consolidate the current state of knowledge on IoT cybersecurity vulnerabilities, drawing from a bibliometric analysis of scholarly publications up to the end of 2023. The bibliometric approach applies mathematical and statistical analyses to track the progression of this research area, focusing on its conceptual, intellectual, and societal structures [20]. It is widely acknowledged for its capability to encapsulate knowledge and record the existing research within a specific field through the use of quantitative metrics, typically across a certain timeframe. The use of bibliometric analysis is prevalent in literature reviews across numerous fields, including innovation, marketing, computer science, and tourism and hospitality management [21–23]. This approach offers an exceedingly objective, structured, and thorough overview of this rapidly expanding area of research. Scopus was selected for this review due to its wide recognition as a highly credible and reliable bibliometric database, boasting the most extensive database of peer-reviewed research [24]. Additionally, its capacity to introduce a broad range of descriptive information, like citation information and authors' affiliations, further confirms its suitability as the main data source for this review. Next subsections will describe the search strategy, data collection, data cleaning and harmonization, and data analysis and visualization tools.

2.1 Search Strategy and Data Collection

The study examined papers up to 2023, as indicated in Fig. 1, to encompass the widest range of literature. The article title was used as the primary search field to retrieve precise and pertinent results on Cybersecurity Vulnerabilities in the Internet of Things (IoT) by using the following query: TITLE ("Internet of Things" OR "IoT" AND "Cybersecurity"). This search included only research papers; therefore, review papers, book reviews, and newspapers have been excluded. The query resulted in

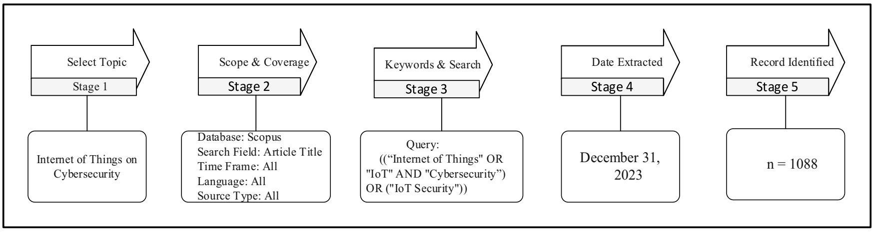


Fig. 1 Flow diagram of the search strategy

1088 articles as shown in the flow diagram of the search strategy in Fig. 1. These articles aim to identify highly cited papers, analyze publication trends and influencing factors over the past decade, determine primary academic platforms for research dissemination, and evaluate common keywords in titles and abstracts of IoT cybersecurity vulnerability papers. The articles were retrieved from the source and organized into an Excel® file for convenient reference.

2.2 Data Cleaning and Harmonization

Cleaning of the data and harmonization are crucial for a reliable bibliometric review. The study utilized OpenRefine and biblioMagika tools [25] to rectify disorderly datasets, including author names, affiliations, and keywords. Initially, Scopus data in.csv format was meticulously selected for cleaning. Clustering tools were then employed to modify specific columns, such as keywords and affiliations. BiblioMagika aided in identifying missing data gaps, facilitating cleaning automatically. After the preliminary phase, keywords were examined for accuracy. During consolidation, multivalued cells were reintroduced with separators. The refined dataset was exported for further analysis, ensuring integrity and reliability.

2.3 Data Analysis and Visualization Tools

After cleaning and harmonizing the data, we conducted frequency and descriptive analyses to identify prolific papers, journals, and authors, as well as the distribution of articles by country and institution. Co-word analysis was then employed to explore the intellectual structure of vulnerabilities in IoT cybersecurity, revealing key themes and trends. BiblioMagika and the Bibliometric R package served as the data analysis and visualization tools, offering comprehensive bibliometric assessments covering various metrics, such as Total Publications (TP), Number of Contributing Authors (NCA), Number of Cited Publications (NCP), Total Citations (TC), Citations Per Paper, Citations Per Cited Paper, Citations Per Author (C/A), Authors Per Paper (A/

P), Citations Per Year (C/Y), Citable Year, h-index, g-index, m-index, and Citation Sum within h-Core. VosViewer was used to identify clusters for key themes and trends in the arena [21, 26–28]. These tools and techniques facilitated a thorough and robust analysis of vulnerabilities in IoT cybersecurity.

3 Results

The analysis addressed a number of publications on IoT and Cybersecurity issues. To answer the research questions included the analysis publication trends publications by authors, publications by countries publications by Institutions, and co-authorship of keywords. In addition, it included future directions of studies.

3.1 Documents Profiles

The first research question addresses the current landscape of IoT and cybersecurity research. The main emphasis of this question is to analyze the publications related to subject area, document type, languages, and source type. The document types included conference papers, research articles, reviews, conference reviews, editorials, books, notes, and short surveys. As shown in Table 1, conference papers are the most common document type, representing 51.84%. Followed by research articles and a book chapter, representing 31.99% and 9.09%, respectively. The field's rapid growth and the interdisciplinary nature of IoT and cybersecurity have attracted researchers and practitioners from various backgrounds to share their latest findings and innovations, which reflects the huge number of conference papers on IoT in the last ten years. Conferences also deliver vital networking opportunities, enabling collaboration and the exchange of knowledge to advance theoretical understanding and practical applications of IoT and cybersecurity.

Table 2 shows five different source types, which are conference proceedings, journals, book series, books, and trade journals. Conference proceedings are the most common source type, representing 43.84% of the total publications. Journals are the second most common source type, with 36.49%, followed by book series at 13.24%. The present study also provides analysis based on the subject area, which provides a vision into the interdisciplinary nature of IoT and cybersecurity research.

3.2 Publication Trends

The Fig. 3 and Table 4 depicts a regular upward publication trend from 2013 to 2023; as usual, some bumps occur in between. There has been an increase in publications, from just four papers in 2013 to 245 papers in 2023. Studies on IoT and cybersecurity

Table 1 Document type

Document type	Total publications (TP)	Percentage (%)
Conference paper	564	51.84
Article	348	31.99
Book chapter	98	9.01
Review	44	4.04
Conference review	15	1.38
Editorial	12	1.10
Book	4	0.37
Note	2	0.18
Short survey	1	0.09
Total	1088	100.00

Table 2 Source type

Source type	Total publications (TP)	Percentage (%)
Conference proceeding	477	43.84
Journal	397	36.49
Book series	144	13.24
Book	57	5.24
Trade journal	13	1.19
Total	370	100.00

have seen a sudden spurt since 2017. As citations represent the role of each publication in the academic area, widely cited works usually reflect the thought-provoking nature and excellent quality of the research. There is a visible trend of increasing papers every year with higher citation rates, which proves that the chair's brash statement 'Internet of Things and cybersecurity are hot and important research topics in the area' is correct. Therefore, the main cause for this raise in citation is the constantly increasing awareness of the need to protect IoT devices, as well as the continuous expansion of IoT devices into various facets of life and important infrastructures. This has forced researchers to pursue novel approaches and techniques when it comes to solving cybersecurity issues, hence there is an abundance of incredibly important research, which actively generates a lot of academic interest. Also, exchange and conjunction between research fields and methodologies have improved the relevance and academical value of the findings which also have raised the rates of citation. The number of citations from 2013 to 2019 illustrates that the research is highly attention-grabbing. While this is a rapid rise from 2017 to 2019, there is a steep decline in 2020 and 2021, a slight growth in 2022, and a sharp drop in 2023 (see Table 4 and Fig. 2). Diminution in the number of publications and citations from 2020 to 2023 indicates probable subjects of discussion, including, but not limited to, changes in

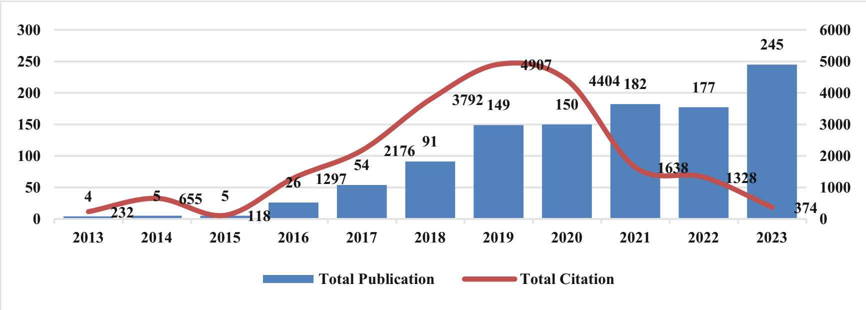


Fig. 2 Total publications and citations by year

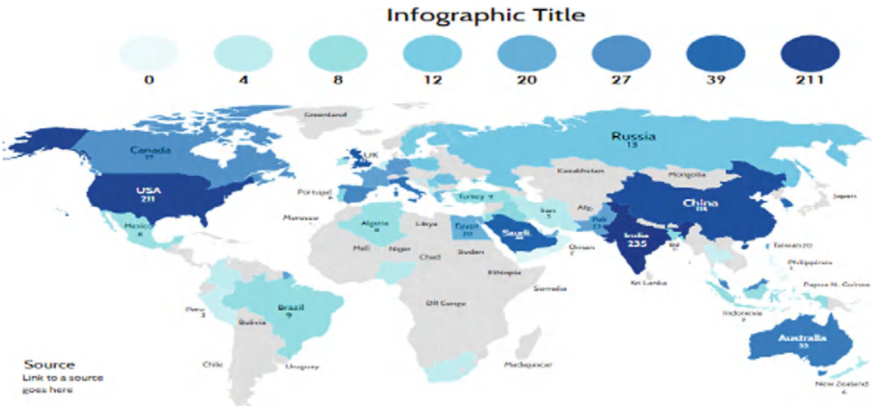


Fig. 3 Worldwide scientific production indexed by Scopus on labour relations

research, external events that shifted the priority lists of research, and saturation of topics in the IoT and cybersecurity sectors (Table 3).

3.3 *Publications by Authors*

Authors belong to publications that refer to books, articles, and research papers. Lee, J. from the Hong Kong University of Science and Technology as one of the most productive authors in the fields of IoT and Cybersecurity with the total publications of 7 and the total citation of 80 is notable. Lee, J. has a notable effect with a citation per publication ratio of 11.43, indicating substantial impact relative to the total number of publications. This author's h-index is 4, showing a consistent citation effect across their research work. A g-index of 7 further supported their impact, reflecting the cumulative impact of their top publications. Similar to McGlone et al., Shan S.N.,

Table 3 Year of publication

Year	TP	NCP	TC	C/P	C/CP	*h*	*g*
2013	4	16	232	58.00	77.33	3	4
2014	5	26	655	131.00	131.00	4	5
2015	5	15	118	23.60	23.60	3	5
2016	26	80	1297	49.88	54.04	12	26
2017	54	171	2176	40.30	43.52	17	46
2018	91	301	3792	41.67	49.25	24	61
2019	149	498	4907	32.93	38.64	26	68
2020	150	530	4404	29.36	34.41	30	64
2021	182	658	1638	9.00	11.78	23	35
2022	177	694	1328	7.50	11.35	16	33
2023	245	850	374	1.53	4.30	10	14
Total	1088						

Note TP = total number of publications; NCP = number of cited publications; TC = total citations; C/P = average citations per publication; C/CP = average citations per cited publication; h = h-index and g = g-index

Table 4 Most productive authors

Author's name	Affiliation	TP	NCP	TC	C/P	C/CP
Lee, J	Ajou University	7	6	80	11.43	13.3
Lee, Y	SKKU College of Information and Communication Engineer	7	6	80	11.43	13.3
Matheu, S.N	Universidad de Murcia	6	5	156	26.00	31.2
Okada, Y	Kyushu University	6	6	36	6.00	6.0
Shi, W	Kyushu University	6	6	44	7.33	7.3
Skarmeta,A	Universidad de Murcia	6	4	92	18.40	23.0
Bose, R	Indian Institute of Technology	5	5	39	7.80	7.8
Kulshrestha, S	Indian Institute of Technology Delhi	5	5	39	7.80	7.8
Abie, H	Norwegian Computing Centre	4	4	18	4.50	4.5
Andrade, R.O	Escuela Politécnica Naciona	4	3	7	1.75	2.3
Baldini, G	European Commission Joint Research Centre	4	3	8	2.00	2.7
Haga, A	Kyushu University	4	4	43	10.75	10.8
Jayawardena, C	Sri Lanka Institute of Information Technology	4	3	68	17.00	22.7
Zhang, Minzhi	Zhongyuan University of Technology	4	4	18	4.50	4.5
Li, Hong	Institute of Information Engineering	4	4	12	3.00	3.0

a researcher at the University of Hong Kong, has 6 research publications and 156 citations. Mathieu, S.N.'s opportunity to shine is evidenced by this impressive citation per publication ratio of 26.00, demonstrating their high-impact articles within the subject area. The h-index and g-index show that Matheu, S.N. has notably contributed to the field through several of her works, which have been cited extensively. Besides, the author Shi, W., of the University of Hong Kong is also noteworthy because of her co-authorship through 6 publications with a cumulative citation count of 44. Although Shi, W. may be slightly less influential in terms of Lee, J., and Matheu, S.N., their citation per paper ratio is 7.33. Still, such researchers exert a considerable influence on the invention of IoT and cybersecurity. Marshaling their joint efforts together, these authors' assistance creates the discourse. They can be credited with the intellectual growth within these varied disciplines, affording them a central role in research circles, see Table 5.

Table 5 Top 20 Countries contributed to the publications

Country	TP	NCP	TC	C/P	C/CP	*h*	*g*
India	235	3065	133	13	23.0	19	55
United States	211	6094	165	29	36.9	38	78
China	115	2184	85	19	25.7	23	46
United Kingdom	62	1278	51	21	25.1	18	35
Saudi Arabia	56	1067	48	19	22.2	16	32
South Korea	39	647	33	17	19.6	14	25
Italy	34	651	30	19	21.7	13	25
Australia	33	1909	26	58	73.4	10	33
Malaysia	29	1653	27	57	61.2	12	29
Spain	29	604	21	21	28.8	13	24
Japan	28	154	21	6	7.3	8	12
Canada	27	1464	23	54	63.7	11	27
Germany	24	124	13	5	9.5	6	11
France	23	818	21	36	39.0	8	23
Pakistan	23	2611	21	114	124.3	11	23
UAE	21	2504	16	119	156.5	9	21
Egypt	20	206	14	10	14.7	6	14
Taiwan	20	812	17	41	47.8	9	20
Morocco	19	167	13	9	12.8	7	12
Greece	18	182	14	10	13.0	7	13

3.4 Publications by Countries

Table 5 and Fig. 3 show the country's most productive in IoT and cybersecurity research, with the United States, China, India, and the UK as the leading countries. The U.S. tops in terms of total publication (TP) and total citation (TC), signifying a massive research volume and a noticeable impact, which illustrates US knowledge's significance and uniqueness. The high TC of 6094 further shows the strong effect of US investigation in the global scientific community. Furthermore, the high h-index also makes this country top-notch in research on IoT and cybersecurity. India keeps its high TP (235), which testifies to the significance of India's contribution. However, the citation numbers as a measure of the impact have less significance as they constitute a few compared to the US. The UK is important among the countries, showing productivity and impact in IoT and cyber security research with a high (1278) NCP number and 62 papers. The UK researchers have an average of solid publications, citation figures, and respectable impact ratings. Likewise, their h-index is good. These four countries cumulatively stand at the center of the investigations on IoT and cybersecurity, where they are making progress in the area and having a substantial influence. Other countries have significant contributions in terms of publications, such as Saudi Arabia (56), South Korea (39), Italy (34), Australia (33), and Malaysia (29).

3.5 Publications by Institutions

Table 6 shows that several institutions stand out among the most advanced institutions in IoT and cybersecurity research for their best performance on different metrics. Although King Abdulaziz University in Saudi Arabia is the most productive, with the maximum arithmetic mean of total publications (21), Amity University in India stands out as the most productive institution regarding the number of total citations (1384).

They held h-indexes of 13 and 14, respectively, reflecting the substantial quality of their research impact. This outstanding effort may be due to King Abdulaziz University and Amity University's strong research orientation, widespread partnerships with industry and academia, and high investments in class-leading infrastructure. The University of Technology Malaysia has experienced a respectable increase in the output of research related to IoT devices and cyber safety. While its total product of 14 may be comparatively lower than Amity University's, it highlights an impressive total citation (11) level. In addition, it has higher C/P (39) and C/CP (49.2) ratios than King Abdulaziz University's C/P (8) and C/CP (10.4). Alongside other institutions, Sungkyunkwan University also has a strong impact on the levels of TP (13) and TC (12), which emphasizes its strong position in this research field. The high impact and quality of its research results are reflected in its h-index, which is 13. Sungkyunkwan University, the University of Chinese Academy of Sciences, the University of Murcia,

Table 6 Most productive institutions with minimum of five publications

Affiliation	Country	TP	NCP	TC	C/P	C/CP	*h*	*g*
King Abdulaziz University	Saudi Arabia	21	177	17	8	10.4	6	13
Amity University	India	14	1384	8	99	173.0	5	14
Universiti Teknologi Malaysia	Malaysia	14	541	11	39	49.2	7	14
Sungkyunkwan University	South Korea	13	236	12	18	19.7	7	13
Beijing University of Technology	China	12	130	8	11	16.3	5	11
Ningbo University	South Korea	12	97	10	8	9.7	4	9
Universidad de Las Américas	Chile	12	41	7	3	5.9	4	6
SRM Institute of Science and Technology	India	12	55	9	5	6.1	4	7
SRM TRP Engineering College	India	12	316	9	26	35.1	6	12
University of Chinese Academy of Sciences	China	12	152	9	13	16.9	6	12
Prince Sattam Bin Abdulaziz	Saudi Arabia	12	89	8	7	11.1	6	9
University of Murcia	Spain	11	266	10	24	26.6	7	11
Vellore Institute of Technology	India	11	80	5	7	16.0	3	8
University of Georgia	United States	10	158	7	16	22.6	5	10
Chitkara University Institute	India	10	36	3	4	12.0	2	6

and Vellore Institute of Technology also demonstrate notable productivity and impact in the field. They leverage research focus, partnerships, and resource investment to contribute meaningfully to IoT and cybersecurity research. These institutions collectively play a crucial role in advancing knowledge and addressing challenges in these domains. Furthermore, other institutions such as Beijing University of Technology, Ningbo University, the University of Murcia, and the University of Georgia demonstrate notable productivity and impact, with TPs of 12, 12, 11, and 10, respectively. While some institutions share the same TP of publications, they show varied investigate influences. For example, SRM TRP Engineering College and the University of Chinese Academy of Sciences have the same total of publications (12), but they have different numbers of NCPs, which are (316), and (152), respectively. Table 9 offers data about essential institutional contributions regarding IoT and cybersecurity research. This data delivers valued visions for future official collaboration and planning of strategic, and it is important for understanding the landscape of present studies about IoT.

3.6 Co-Authorship by Analysis

Co-authorship analysis is concerned with identifying authors' dyadic affiliation, which pertains to their interaction in published documents. They reveal the patterns of collaboration, the formal organization of individuals involved in research work, and the impact of collaboration on the resultant advancement in knowledge. Its application is crucial for comprehending the extant and potential interaction within research communities and optimizing collaborative efforts.

3.6.1 Co-Authorship by Author

Derived from the analysis of correlative terms in earlier research, co-occurrence networks provide detailed perspectives on the complex interactions within different fields of study. Co-occurrence analysis generates these networks as a powerful structural representation that unfolds critical insights into theoretical structures that underlie particular domains of inquiry. The beneficiaries of such information are not limited to the academia but also professionals, policy makers and scholars who are deeply engrossed with the subject. In Fig. 4, we reveal a co-authorship network involving 32 nodes that represent individual authors in the research landscape. Here, each node stands for an author and the edges between them trace the co-authorship relationships.

A detailed look reveals the appearance of separate clusters, each with its own peculiarities and details. The red cluster becomes the main source of influence in the collaborative network, consisting of more than six authors. Grouped in co-authorship chains of extremely high production, these writers represent the core of the most influential co-authorship network. The blue and green clusters, which all have networks covering more than five authors, represent big collaborative endeavors that add much to the overall research discourse. Such a graphic model not only solidifies the present collaborative dynamics but also points to the potential of developing more powerful relationships among researchers. In this way, the joint efforts of authors, policymakers, and scholars will move the IoT-Cybersecurity research landscape into new dimensions of innovation and knowledge sharing.

3.6.2 Co-Authorship by Countries

Co-occurrence networks are derived from the analysis of correlative countries in earlier research and offer detailed perspectives on the intricate interactions in various research areas. Figure 5 shows a country network comprising 22 nodes that stand for

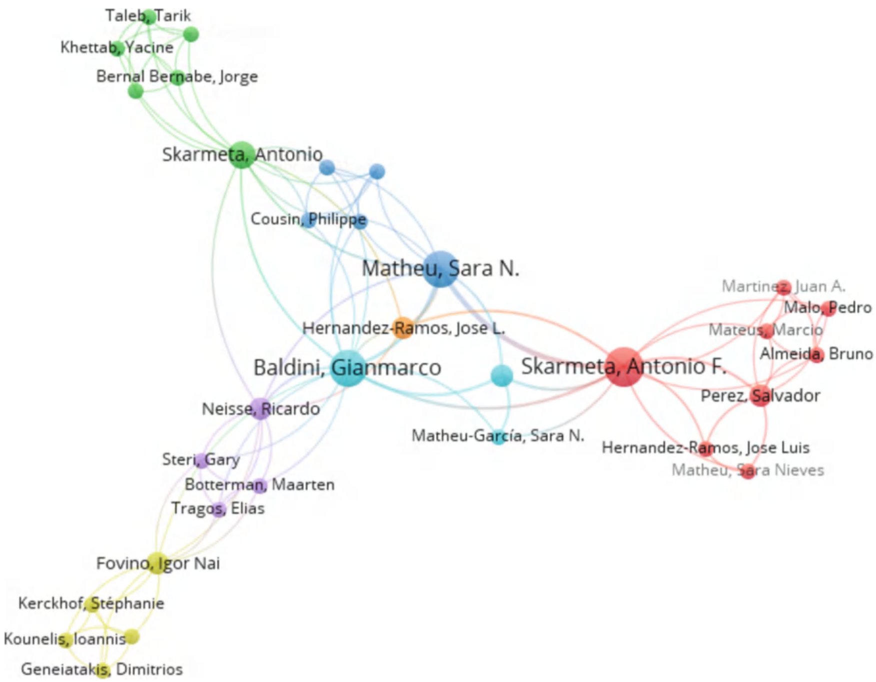


Fig. 4 Network visualisation map of the co-authorship by authors

countries in the research landscape. Each node represents an author, and the edges connecting them represent the co-authorship relationships. The close study shows individual groups, each of their own features and characteristics. The US and India cluster in the collaborative network becomes the main source of influence. Clustered in exceedingly high–production countries chains, these authors form the heart of the most powerful countries network. The UK, Saudi Arabia, and Germany clusters are major collaborative efforts that generate much scholarly collaboration. Some groups from the side-lines come up to smaller collaborative units. Despite the fact that such clusters are weaker in collaboration than their more extensive counterparts, they are an integral part of the larger network. Such connectivity opens a landscape for development, offering a chance for improved overall country collaboration in IoT-Cybersecurity research.

3.7 *Co-Occurrence Analysis*

Co-occurrence analysis aims at establishing the relationship between the frequency of a particular term or concept with that of other related terms or concepts in a given

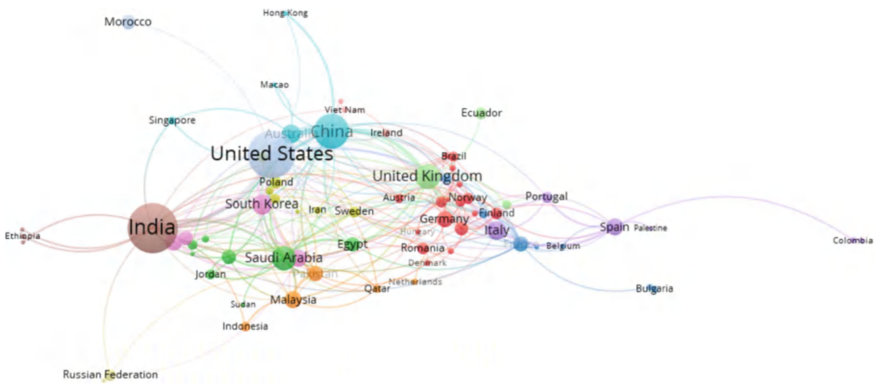


Fig. 5 Network visualisation map of the co-authorship by countries

document. One can use it to discover relationships, trends, and structure within a topic by seeing which signifiers most commonly co-occur.

Co-occurrence networks are built upon the frequent terms detected in earlier scholarly studies (see Fig. 6). Co-occurrence analysis allows the exploration of the network structure, the understanding of which can give important information on the theoretical background of a research domain. Such insights provide benefits to the practitioners, policy makers, and scholars in the field of concern, but also for understanding interrelations among terms and concepts. Figure 6 shows co-occurrence networks that help to comprehend commonly used terms and their relationships in the research field. The co-occurrence network functions as a visual illustration of relationships between topics detected by the available literature [29, 30]. The structure is based on connected lines and nodes, with the larger nodes representing the dominant themes evident in the literature. These network structures provide policymakers and researchers with the chance to use them in order to simplify the reorganization of the current information and research findings. In the area of IoT Cybersecurity, the terms of the IOT and security have come up as prominent concepts used by previous researchers making them quite significant within this domain [31, 32]. Table 7 is shown all themes.

Table 8 offers a sort of research or literature review on different themes of cybersecurity and the Internet of Things (IoT) [33–35]. The key components of the table are listed below: The first column presents keywords or phrases related to each theme or cluster. Therefore, these keywords are probably an indication of the point of interest of the research or the literature that is being reviewed. The thematic cluster or category to which each keyword belongs is indicated in the second column. For instance, "Internet of Things" and "Information Security" can be found in Cluster 1, "IoT Security" in Cluster 2, and so on. The third column includes numerical data that reflects the total link strength of each keyword. Link strength is also a way of describing the relationship of a particular keyword with the rest of the cluster or theme. The fourth column represents the frequency of occurrences of each keyword in the literature or

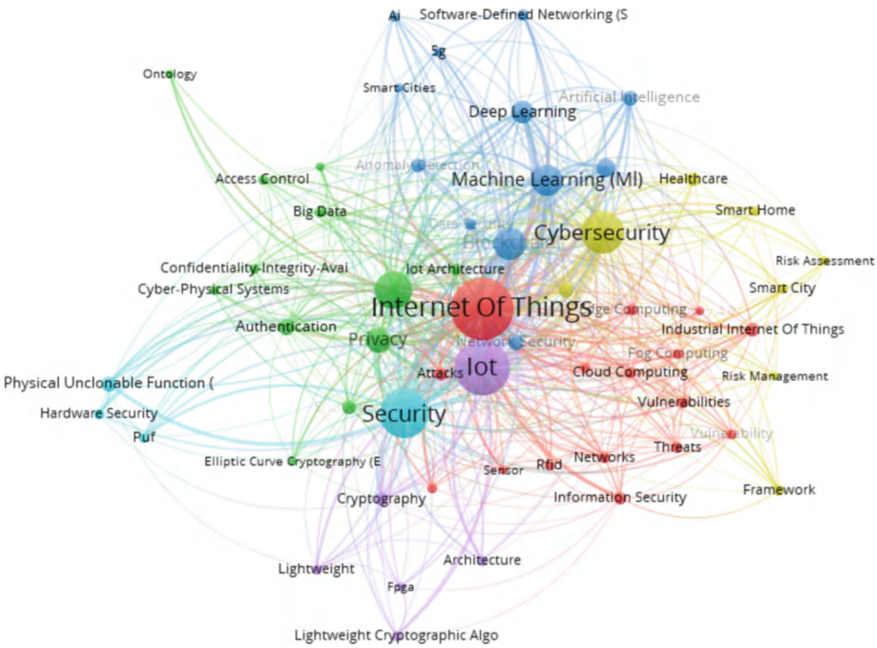


Fig. 6 Network visualisation of the author's keywords

research being reviewed. The final column provides a short description or name of each theme or cluster. The themes are the areas of focus within the cybersecurity and IoT domains.

Theme 1: The emphasis is on the Internet of Things (IoT) with the keywords associated with information security, attacks, vulnerabilities, networks, and issues. Also, the theme is focused on IoT with the key terms including "Internet of Things", "Information Security" and "Attacks." The total strength of the link is highest for the "Internet of Things," showing how crucial it is in this theme.

Theme 2: The security of IoT, including privacy, authentication, encryption, and other security issues. Also, this aspect represented by the theme deals with the security of IoT, and some keywords associated with security include "IoT Security," "Privacy," and "Authentication." The link strength of "privacy" is high, revealing its essential role in IoT security.

Theme 3: Highlights machine learning and artificial intelligence to include deep learning, intrusion detection systems, and network security. Also, focuses on the intersection of IoT and machine learning/artificial intelligence, with keywords including "Machine Learning," "Deep Learning," and "Artificial Intelligence." Deep Learning is the one with the highest link strength, and this shows its importance within this theme.

Theme 4: Covers cyber security challenges such as cyber-attacks, health care security, risk assessment, and risk management. In addition, "Cybersecurity" has

Table 7 IoT cybersecurity themes

Author's keywords	Cluster	Link strength	Occurrences	Theme
Internet of things	1	342	393	Theme 1:
Information security	1	17	18	
Attacks	1	16	17	
Industrial internet of things	1	16	19	
Edge computing	1	15	15	
Vulnerabilities	1	15	15	
Networks	1	14	14	
Challenges	1	11	11	
Threats	1	11	13	
Cloud computing	1	10	11	
Cyber threats	1	9	10	
Iot security	2	101	144	Theme 2:
Privacy	2	62	62	
Authentication	2	26	26	
Encryption	2	19	20	
Big data	2	15	15	
Iot architecture	2	14	14	
Access control	2	11	11	
Confidentiality-integrity-availability (Cia) triad	2	10	10	
Cyber-physical systems	2	10	10	
Elliptic curve cryptography	2	8	8	
Security and privacy	2	8	8	
Machine learning (Ml)	3	90	97	Theme 3:
Deep learning	3	51	51	
Intrusion detection system (Ids)	3	36	37	
Artificial intelligence	3	24	24	
Network security	3	23	23	
Anomaly detection	3	19	21	
Software-defined networking	3	15	15	
Data security	3	12	13	
Ai	3	10	11	
Cybersecurity	4	175	182	Theme 4:
Cyberattack	4	24	25	
Healthcare	4	19	19	
Framework	4	13	13	
Smart city	4	11	11	

(continued)

Table 7 (continued)

Author's keywords	Cluster	Link strength	Occurrences	Theme
Smart home	4	10	11	
Risk assessment	4	9	9	
Risk management	4	9	9	
Iot	5	267	296	Theme 5:
Cryptography	5	20	20	
Architecture	5	11	11	
Lightweight	5	10	10	
Lightweight cryptographic algor	5	10	10	
Security	6	229	233	Theme 6:
Physical unclonable function	6	20	23	
Hardware security	6	12	12	
Ontology	6	8	8	
Blockchain	6	102	104	

significant link strength which highlights its essential status in this cybersecurity topic.

Theme 5: Focuses on subjects related to the Internet of Things, such as cryptography, architecture, and lightweight cryptographic algorithms. Besides, "Cryptography" has significant link strength, indicating the importance of securing IoT-related systems.

Theme 6: It is dedicated to security issues, such as physical unclonable functions (PUF), hardware security, ontology, and blockchain. Moreover, "Blockchain" is characterized by a high link strength, which signifies its importance and connectedness in this security topic. Table 8 gives a detailed overview of the multidimensional and interlinked research in the field of IoT and cybersecurity. This table allows researchers as well as readers to appreciate the dominance of certain keywords, their interrelations, and the emphasis within each of them.

4 Future Directions: Discussion and Findings

In this study, we collected and analysed articles in the field of cybersecurity related to an IoT. From these studies, it can be seen that extensive research has been conducted on this topic, and the number of publications is increasing. As analysed in section (Sect. 3. 2) There has been a noted increase for articles under the keywords of IoT ad cybersecurity from the onset of the year 2013 and up to the year 2023. The figure from four papers in 2013 to 245 papers in 2023 suggests an expanding body of literature and both theoretical and applied research focused on these areas. Even more important is the fact that the rates of citation also rose, which proves the significance and influence of these works. Thus, the future research concerning the IoT and cybersecurity needs to be continued due to the emerging new problems and opportunities, which make this field interesting for further studies. From the analysis of publications by countries (Sect. 3.5) the five top countries that develop IoT and cybersecurity technologies are the United States, China, India, and the United Kingdom. The United States is quite indicative of its effect, with the highest total citations, which would point towards the contribution of America's research to the global scene. Similar to what has been established earlier, the research activities of these countries are underpinned by sound policies as regards to IoT and cybersecurity innovation. Future research should focus on this area and try to determine the effects of such policies as a way of understanding how policy can support technology development in these fields [36].

In the context of interdisciplinary collaboration, an IoT and cybersecurity research focus areas are greatly interrelated, using various branches of knowledge such as computer engineering, informational and technical science. The conclusion drawn from the observations in (Sect. 3.7) that there exists a significant connection between co-occurrence networks of the author keywords, whether it is between machine learning and artificial intelligence or blockchain. Further research studies should encourage collaboration among different specialties to advance knowledge in IoT and cybersecurity as well as solve new sophisticated issues. This is bound to encourage the development of more complete and effective solutions to the given problems. In addition, the findings presented in mentioned section presented several technologies such as edge computing, fog computing, and blockchain that are most relevant in today's matters concerning IoT and cybersecurity. More research should be directed on how the technologies mentioned herein can be combined, to provide the necessary security, for seamless running of IoT systems. Future investigations focusing on the use of these technologies in different industries, including, healthcare, smart cities and industry IoT will be useful. Moreover, the energy efficiency and sustainability represent another topic which is a trend emerging from the co-occurrence analysis of the IoT and cybersecurity literature. The investigation in the subsequent studies needs to be directed at examining how IoT systems should be implemented in order to minimize or possibly avoid negative impact on the environment. This also entails researching novel behaviours, groups, and apparatuses that reduce power usage while still being safe and efficient. Besides, the smart logistic is another potential area of research for the future, which can be derived from the key terms analysis. Today, IoT technologies can help make such an impact that would imply the increase of logistics efficiency due to improved data control and real-time visualization of processes. As a result, future research should consider exploring the IoT as a means of implementing and expanding logistics and supply chain applications toward achieving these benefits. Furthermore, the literature review reveals the absence of prior research on the system, mechanism, law, and policies related to IoT and cybersecurity. Referring to the publication trends and the distribution geography analysed in this paper, it is necessary to mention that the establishment of sound policies and regulations is critical to the development of these areas. Future scholars must fill the gap by researching on the outcome of the existing policies and policies that may be developed in order to foster innovation while achieving security and privacy objectives.

When summarizing the existing literature, we realized there is a lack of research on the explores cases of smart-port policies regarding IoT, cybersecurity, and their economy and encourage intersectoral collaboration. About edge computing and blockchain, the basics of IoT systems' energy efficiency, logistics optimization, and secure IoT deployment policy frameworks comprehension.

5 Conclusion

The Internet of Things and cybersecurity is considered a revolutionary field in a world dominated by automation. The Internet of Things and Cyber Security is the future of Internet-based businesses. Internet of Things technologies have taken over most aspects of life and have surpassed the refinements of artificial intelligence. Smart devices have contributed to providing support to most sectors. However, security vulnerabilities and continuous device hacks have increased concerns about violating privacy and security. Therefore, security vulnerabilities correspond to the emergence of the Internet of Things. Practitioners and academics have increasingly called on researchers to explore cybersecurity's relationship to the Internet of Things. The current study complements previous literature by taking an insight into and exploring the research stream related to cybersecurity and the Internet of Things. This study aims to summarize and reorganize the results of previous literature related to cybersecurity and the Internet of Things to keep pace with the massive research stream in this field. Previous literature recommended many programs to conduct bibliometric study, but previous literature claimed that VosViewer and R-Studio are the programs that give the most reliable results. The current study shed light on the current research, most years and relevant sources, and focused on reorganizing the literature results by presenting a network of main keywords in the titles and keywords. Moreover, we identified important topics in cybersecurity and opened promising doors for future studies. The United States, Australia, and the United Kingdom are the largest contributors to cybersecurity and the Internet of Things. This result is considered encouraging for academics to increase scientific cooperation with these countries. Governments can benefit from countries contributing to this field to develop wise regulations and policies that mitigate and limit privacy violations by Internet of Things devices. Governmental institutions can benefit from filtering results for keywords and incorporating important clarifications in the field of security and privacy development. Focusing on the areas explored in this study and developing research projects related to the topics identified by universities would provide many solutions to the technical industry and provide a secure, automated world.

References

1. Alwahedi, F., Aldhaheri, A., Ferrag, M.A., Battah, A., Tihanyi, N.: Machine learning techniques for IoT security: Current research and future vision with generative AI and large language models. Internet Things Cyber-Physical Syst. (2024).
2. Sharma, N., Shamkuwar, M., Singh, I.: The history, present and future with IoT. Internet things big data Anal. smart Gener. 27 (2019).
3. Fizza, K., Banerjee, A., Jayaraman, P.P., Auluck, N., Ranjan, R., Mitra, K., Georgakopoulos, D.: A survey on evaluating the quality of autonomic internet of things applications. IEEE Commun. Surv. Tutorials. (2022).
4. Quy, V.K., Nguyen, D.C., Van Anh, D., Quy, N.M.: Federated Learning for Green and Sustainable 6G IIoT Applications. Internet of Things. 101061 (2024).
5. Amiri, A., Tavana, M., Arman, H.: An Integrated Fuzzy Analytic Network Process and Fuzzy Regression Method for Bitcoin Price Prediction. Internet of Things (Netherlands). **25**, 101027 (2024).
6. Alomari, A., Kumar, S.A.P.: Internet of Things Securing IoT systems in a post-quantum environment : Vulnerabilities , attacks , and possible solutions. Internet of Things. **25**, 101132 (2024).
7. Kaur, B., Dadkhah, S., Shoeleh, F., Neto, E.C.P., Xiong, P., Iqbal, S., Lamontagne, P., Ray, S., Ghorbani, A.A.: Internet of things (IoT) security dataset evolution: Challenges and future directions. Internet of Things. 100780 (2023).
8. Al-hchaimi, A.A.J., Alomari, M.F., Muhsen, Y.R., Sulaiman, N. Bin, Ali, S.H.: Explainable Machine Learning for Real-Time Payment Fraud Detection: Building Trustworthy Models to Protect Financial Transactions. In: *International Conference on Explainable Artificial Intelligence in the Digital Sustainability*. Springer (2024).
9. Muhsen, Y.R., Al-hchaimi, A.A.J.: Modelling Intelligent Agriculture Decision Support Tools to Boost Sustainable Digitalization: Evidence from MCDM Methods. In: *International Conference on Explainable Artificial Intelligence in the Digital Sustainability*. Springer (2024).
10. Hahn, C., Kim, J., Kwon, H., Hur, J.: Efficient iot management with resilience to unauthorized access to cloud storage. IEEE Trans. Cloud Comput. **10**, 1008 (2020).
11. Krishna, C.G.L., Murphy, R.R.: A review on cybersecurity vulnerabilities for unmanned aerial vehicles. In: *2017 IEEE International Symposium on Safety, Security and Rescue Robotics (SSRR)*. IEEE (2017).
12. Dadkhah, M., Lagzian, M., Rahimnia, F., Kimiafar, K.: What do Publications say about the Internet of Things Challenges/Barriers to uninformed Authors? A bibliometric Analysis. JLIS. it. **11**, 77 (2020).
13. Dadkhah, M., Lagzian, M., Santoro, G., Drobics, M.: A brief overview on status of Internet of Things research in Iran. Int. J. Internet Things Cyber-Assurance. **1**, 232 (2020).
14. MacDonald, K.I., Dressler, V.: Using citation analysis to identify research fronts: A case study with the Internet of Things. Sci. Technol. Libr. **37**, 171 (2018).
15. Mishra, D., Gunasekaran, A., Childe, S.J., Papadopoulos, T., Dubey, R., Wamba, S.: Vision, applications and future challenges of Internet of Things: A bibliometric study of the recent literature. Ind. Manag. Data Syst. **116**, 1331 (2016).
16. Mora, L., Bolici, R., Deakin, M.: The first two decades of smart-city research: A bibliometric analysis. J. Urban Technol. **24**, 3 (2017).
17. Muhuri, P.K., Shukla, A.K., Abraham, A.: Industry 4.0: A bibliometric analysis and detailed overview. Eng. Appl. Artif. Intell. **78**, 218 (2019).
18. Rejeb, A., Abdollahi, A., Rejeb, K., Treiblmaier, H.: Drones in agriculture: A review and bibliometric analysis. Comput. Electron. Agric. **198**, 107017 (2022).
19. Adunlin, G., Diaby, V., Xiao, H.: Application of multicriteria decision analysis in health care: a systematic review and bibliometric analysis. Heal. Expect. **18**, 1894 (2015).
20. Zupic, I., Čater, T.: Bibliometric methods in management and organization. Organ. Res. methods. **18**, 429 (2015).

21. Al-Enzi, S.H.Z., Abbas, S., Abbood, A.A., Muhsen, Y.R., Al-Hchaimi, A.A.J., Almosawi, Z.: Exploring Research Trends of Metaverse: A Bibliometric Analysis. In: *International Multi-Disciplinary Conference-Integrated Sciences and Technologies*. Springer (2023).
22. Iqbal, A., Ramachandran, S., Siow, M.L., Subramaniam, T., Mohammad Afandi, S.H.: Meaningful community participation for effective development of sustainable tourism: Bibliometric analysis towards a quintuple helix model. J. Outdoor Recreat. Tour. **39**, 100523 (2022).
23. Song, G., Wu, J., Wang, S.: Text mining in management research: a bibliometric analysis. Secur. Commun. Networks. **2021**, 1 (2021).
24. Feng, Y., Zhu, Q., Lai, K.-H.: Corporate social responsibility for supply chain management: A literature review and bibliometric analysis. J. Clean. Prod. **158**, 296 (2017).
25. Ahmi, A.: OpenRefine: An approachable tool for cleaning and harmonizing bibliographical data. In: *AIP Conference Proceedings*. AIP Publishing (2023).
26. Muhsen, Y.R., Husin, N.A., Zolkepli, M.B., Manshor, N.: A systematic literature review of fuzzy-weighted zero-inconsistency and fuzzy-decision-by-opinion-score-methods: assessment of the past to inform the future. J. Intell. Fuzzy Syst. 1 (2023).
27. Khairan, H.E., Zubaidi, S.L., Muhsen, Y.R., Al-Ansari, N.: Parameter Optimisation-Based Hybrid Reference Evapotranspiration Prediction Models: A Systematic Review of Current Implementations and Future Research Directions. Atmosphere (Basel). **14**, 77 (2022).
28. Kareem, B.A., Zubaidi, S.L., Al-Ansari, N., Muhsen, Y.R.: Review of Recent Trends in the Hybridisation of Preprocessing-Based and Parameter Optimisation-Based Hybrid Models to Forecast Univariate Streamflow.
29. Muhsen, Y.R., Husin, N.A., Zolkepli, M.B., Manshor, N., Al-Hchaimi, A.A.J.: Evaluation of the Routing Algorithms for NoC-Based MPSoC: A Fuzzy Multi-Criteria Decision-Making Approach. IEEE Access. (2023).
30. Al-Enzi, S.H.Z., Abbas, S., Abbood, A.A., Muhsen, Y.R., Al-Hchaimi, A.A.J., Almosawi, Z.: Exploring Research Trends of Metaverse: A Bibliometric Analysis BT - Beyond Reality: Navigating the Power of Metaverse and Its Applications. Presented at the (2023).
31. Al-Hchaimi, A.A.J., Sulaiman, N. Bin, Mustafa, M.A. Bin, Mohtar, M.N. Bin, Mohd, S.L.B., Muhsen, Y.R.: Evaluation Approach for Efficient Countermeasure Techniques Against Denial-of-Service Attack on MPSoC-based IoT Using Multi-Criteria Decision-Making. IEEE Access. (2022).
32. Al-Hchaimi, A.A.J., Sulaiman, N. Bin, Mustafa, M.A. Bin, Mohtar, M.N. Bin, Mohd Hassan, S.L.B., Muhsen, Y.R.: A comprehensive evaluation approach for efficient countermeasure techniques against timing side-channel attack on MPSoC-based IoT using multi-criteria decision-making methods. Egypt. Informatics J. **24**, 351 (2023).
33. Zaidan, A.S., Alshammary, K.M., Khaw, K.W., Yousif, M., Chew, X.: Investigating Behavior of Using Metaverse by Integrating UTAUT2 and Self-efficacy. In: *International Multi-Disciplinary Conference-Integrated Sciences and Technologies*. Springer (2023).
34. Alhasnawi, M.Y., Said, R.M., Daud, Z.M., Muhammad, H.: Enhancing managerial performance through budget participation: Insights from a two-stage A PLS-SEM and artificial neural network approach (ANN). J. Open Innov. Technol. Mark. Complex. **9**, 100161 (2023).
35. Alhasnawi, M.Y., Alshdaifat, S.M., Aziz, N.H.A., Almasoodi, M.F.: Artificial Intelligence and Environmental, Social and Governance: A Bibliometric Analysis Review. In: *International Conference on Explainable Artificial Intelligence in the Digital Sustainability*. Springer (2024).
36. Husin, N.A., Zolkepli, M.B., Manshor, N., Al-Hchaimi, A.A.J., Albahri, A.S.: Routing Techniques in Network-On-Chip Based Multiprocessor-System-on-Chip for IOT: A Systematic Review. Iraqi J. Comput. Sci. Math. **5**, 181 (2024).

Artificial Intelligence Implementation in the Tourism Industry: A Literature Review

Weiming Wang, Noorminshah A. Iahad, and Mohammed A. Al-Sharafi

Abstract Artificial Intelligence (AI) innovation speaks to a sizeable possibility to transform superior encounters and operational efficiencies within the travel, tourism industry. This article offers a literature review that investigates the vital openings and demanding situations associated with AI utilisation in tourism. The opportunities contain enhancement of optimisation of business decisions, personalised marketing message, automation of customer service through chatbots, and enhancement of both indoor and outdoor service experiences. However, those benefits come with challenges, such as technical constraints, financial barriers, legal considerations, and socio-ethical issues. Findings discover that AI can significantly improve efficiency and customer satisfaction. Addressing the challenges is crucial for the successful adoption of AI. Future research is suggested to enhance AI reliability, reduce costs, ensure data privacy, and balance automation with human interaction. These insights aim to direct tourism organisations and technology companies to leverage AI to improve traveller experiences.

Keywords Artificial intelligence · AI · Tourism · Opportunities · Challenges

W. Wang (✉)
Azman Hashim International Business School (AHIBS), Universiti Teknologi Malaysia, Jln Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia
e-mail: weimingwang@ieee.org

N. A. Iahad
Department of Information Systems, Faculty of Management, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

M. A. Al-Sharafi
IRC for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_9

1 Introduction

Artificial Intelligence (AI) is significant transforming various industries, including tourism industry. Insights highlight the increasing impact of AI on the tourism industry. The global market for AI in hospitality and tourism was valued at USD 13.85 billion in 2022. It is projected to grow at a compound annual growth rate (CAGR) of 35.8% during the forecast period, reaching a value of USD 158.40 billion by 2030 [15].

The integration of AI in tourism is revolutionising the industry by enhancing marketing strategies, service provision, and customer insights. AI-driven tools are enhancing operational efficiency and personalisation, contributing to increased customer satisfaction and profitability. For example, AI's ability to deal with and analyse exceptional sums of customer data enables hyper-personalisation, tailoring travel experiences to individual preferences with improved satisfaction [4].

However, the implementation of AI also presents challenges, involving technical issues, financial barriers, and legal and socio-ethical considerations. In order to utilise AI's benefits while mitigating its risks for enterprises, it is vital to tackle with these challenges to assure reliable and complete AI implementation within travel, tourism and hospitality sector.

This study is in the aim to examine both opportunities and challenges of AI in the tourism industry, which provides a comprehensive review of current applications. In the meanwhile, it also identifies the essential areas for future research. Stakeholders can better utilise AI technologies to enhance service delivery and customer experiences by understanding the transformative capability of AI and the associated challenges.

2 Literature Review

2.1 Artificial Intelligence (AI)

Artificial Intelligence (AI) is to a wide range of technologies, making machines to perform tasks just like human beings. It is widely accepted that the AI technologies include machine learning, robotization, deep learning, and big data, each contributing uniquely to the capabilities and implementations of AI [31].

Machine learning (ML) is a subset of AI that centres on developing algorithms that empower computers to learn from and make predictions based on previous data. This innovation is vital in examining consumer behaviour, proposing marketing strategies, detecting fraud, and generating informed statistics [17].

ML algorithms can deal with vast amounts of data to identify tendencies and patterns, directing product or service providers to optimise their services, tailor marketing efforts, and enhance customer experiences. Robotization refers to the use of robots to carry out tasks usually executed by humans. The advantages of robotization are significant in terms of efficiency and productivity. Norfleet [41] pointed out that the productivity of a robot in one minute can be equivalent to a human's productivity in fifteen minutes. Robots are increasingly used for customer service, cleaning, and guiding tours, offering reliable, high-quality service while reducing operational costs [44].

Deep learning, a specific area of machine learning, uses layered neural networks, called deep neural networks (DNNs), to simulate how the human brain makes decisions [23]. This technology drives many applications that improve automation and decision-making. Examples include digital assistants like Siri and Alexa, voice-controlled TV remotes, credit card fraud detection, self-driving cars, and generative AI (Gen-AI) that produces text, images, and more [23]. Deep learning helps analyse large datasets, personalise customer interactions, and enhance predictions for better service.

Big data refers to the large volumes of data that can be analysed computationally to reveal patterns, trends, and associations, especially relating to human behaviour and interactions. De Mauro, Greco, and Grimaldi [11] define big data as "information assets characterised by such a high volume, velocity, and variety to require specific technology and analytical methods for its transformation into value." Big data analytics can be used to develop targeted marketing plans, improve service offerings, forecast trends, and enhance decision-making processes. By leveraging big data, organisations can gain deeper insights into customer preferences, optimise resource allocation, and create more personalised experiences.

AI encompasses a broad spectrum of technologies transforming various industries by enhancing operational efficiency, personalising customer experiences, and providing deeper insights through data analytics. The continued integration and advancement of AI technologies promise to drive innovation and growth, creating more dynamic and responsive services that meet the evolving needs of users.

2.2 Tourism

Tourism is a multifaceted industry that includes traveling, accommodation, and recreation activities, making it a significant contributor to the global economy [48]. The industry involves a complex network of stakeholders, including travel agencies, airlines, hotels, tour operators, and various ancillary services, all working together to provide exceptional customer experiences.

The COVID-19 pandemic profoundly impacted the tourism industry, leading to unprecedented disruptions and economic downturns. Despite these challenges, the resilience of tourism sector is evident as travel demand remains strong, particularly in the UK and European countries [13]. Similarly, despite a decline in Americans' confidence of financial well-being, the sustained intent to spend on travel suggests that the sector may be somewhat insulated from a moderate economic downturn in the coming year [12].

Tourism is not only a major economic driver but also a significant source of employment. According to the World Travel and Tourism Council (WTTC), the tourism industry is expected to contribute to the world economy with an average annual growth rate of 5.8%, more than twice the predicted average annual growth rate of 2.7% for the global economy. Over the same time frame, the industry is projected to create 126 million additional jobs, highlighting its pivotal contribution to worldwide employment [53].

The economic benefits of tourism go beyond just direct revenue. It also boosts related sectors like transportation, hospitality, and entertainment, creating a ripple effect that helps local economies. Furthermore, tourism promotes cultural exchange, helps preserve cultural heritage, and supports environmental conservation. It is clear that the global tourism industry is poised to recover and grow significantly in the next few years.

2.3 AI in Tourism

The application of AI innovation holds a significant potential to reshape the tourism businesses by enhancing operational performance and enhancing customer experiences. As previously mentioned, technologies such as machine learning, Robotization, deep learning, and big data are widely utilised in the tourism sectors.

Applications of Machine Learning in Tourism

Machine learning (ML) improves personalised travel experiences and optimises operations within the tourism enterprise. By analysing customer data, ML algorithms assist tourism providers in identifying traveller preferences and behaviours, enabling customised recommendations for destinations, accommodations, and events [42]. For instance, providers and online travel agencies make use of ML to offer customised flight and accommodations suggestions primarily based on a customer's searching history and travel intentions. Hospitality providers use predictive analytics to anticipate traveller needs, providing more tailored services and products.

Traditional techniques regularly rely on small, profoundly dissected datasets, whereas ML leverages large datasets, frequently from user-generated content (UGC), to give broader insights [32]. Mariani et al. [35] highlight the challenge of integrating small, detailed datasets with broad ML insights, emphasising the need for a systematic integration of field expertise with data science. On the other hand, existing ML applications frequently incorporate traditional methodologies in an unsystematic manner. The lack of thorough documentation and justification might lead to incomplete or biased insights.

Applications of Robotization in Tourism

Robotization consists of integrating robots into numerous tasks within the tourism industry. This innovation can upgrade operational performance, performance, and customer satisfaction. Robots are increasingly applied in accommodation units, with the Henn-na Hotel in Japan being an instance of fully robot-operated hotel [49]. At Henn-na Hotel, robots carry out tasks including check-ins, and room service, giving information to visitors, with reduced operational expenses, and consistent service quality.

However, the utilisation of robotization come with challenges. Firstly, robot-driven services are less flexible as compared to human-orientated services. Secondly, adopting robots requires substantial reengineering of service delivery processes, including new procedures, systems of control, and updates to service operations manuals [24]. Moreover, while robots can operate 24/7, enhancing performance and economic productivity, they could break down and require maintenance, which leads to disrupting service and incurring additional costs. Client acceptance is another important factor, not all clients accept to interact with robots. Evidence suggests that while some appreciate the novelty and efficiency of robots, others prefer human interaction and may be willing to pay less for robot-delivered services [25].

Applications of Deep Learning in Tourism

Deep learning (DL) offers improved capabilities in analysing and processing large amounts of data to enhance operational and customer-related aspects in the tourism industry. One primary application of DL in tourism is parameter forecasting. DL models, especially Long Short-Term Memory (LSTM) networks, are extensively applied for predicting traveller demands [14]. For instance, Wu et al. [54] illustrated the use of LSTM models to forecast tourist demands, providing valuable insights that help service providers plan for peak seasons and manage their operations more effectively. Another important application of DL in tourism is sentiment analysis. DL understands and processes online reviews, detecting deceptive reviews and analysing sentiments in social media posts.

Despite the benefits, DL in tourism faces limitations. One major challenge is the explainability of DL models, regularly named "black-box" because of their complicated and non-linear computations, making it difficult to interpret the predictions and decisions [14]. The lack of transparency can restrain trust and acceptance of DL. Ethical and privacy issues also pose substantial obstructions, especially regarding the collection, use, and storage of personal data used in training DL models. The black-box nature of these models can raise concerns about data privacy and security [14].

Applications of Big Data in Tourism

Big facts are fundamental in the tourism enterprise, empowering the coping with huge number of data generated by service providers and clients. This data-driven approach improves decision-making and operational efficiency. One example of big data application in tourism is through UGC. Researchers and practitioners make use of these data to gain customer evaluations and inclinations, which inform service improvements and marketing strategies [45]. For instance, analysing online reviews can help identify common issues and areas for improvement in service delivery.

Despite its benefits, they utilisation of big data in tourism faces challenges. Ensuring data privacy and security is critical as the increasing volume of data collected heightens the risk of data breaches and misuse of personal information [36]. Compliance with data protection regulations and implementation of robust security measures are important to protect consumer data and maintain trust. Another challenge is the combination and analysis of diverse data sources. It is required to use cutting-edge tools and methodologies to combine and analyse successfully since the tourism data come from different platforms and formats. The complexity of managing and interpreting datasets can be a barrier for companies, particularly small enterprises with limited resources [36].

Research Questions

1. ***What are the primary opportunities of implementing AI in the tourism industries?***

This question helps identify how AI can be leveraged to improve operational efficiency and customer experiences in the tourism context. It will investigate advantages such as personalised marketing, optimised resource management, and improved service delivery.

2. ***What are the significant challenges faced by the tourism enterprise in the usage of AI?***

This question seeks to uncover the obstacles that hinder the adoption and implementation of AI in tourism and hospitality. It includes technical challenges, financial barriers, legal considerations, and socio-ethical dilemmas that need to be addressed for successful integration.

3. ***What are the potential future research directions for the development and application of AI in the tourism sector?***

This question outlines the areas where further research is needed to advance the understanding and application of AI in tourism. It can contribute to overcoming existing challenges and maximising the advantages of AI technologies in the tourism industry.

3 Methodology

This study employs a literature review methodology. The selected articles were sourced from Scopus database. Scopus is used for the reason of its reputation as one of the maximum well-acknowledged document databases all over the world [37].

The data for the sample article selection was obtained until May 18, 2024. The data collection procedures used a modified version of the PRISMA protocol, as delineated in Fig. 1 [39]. To do this analysis, the selected publications were confined in the articles that only contained terms associated with "AI" and "tourism" within the titles. This procedure can help ensure the articles obtained were highly relevant to this study's subject.

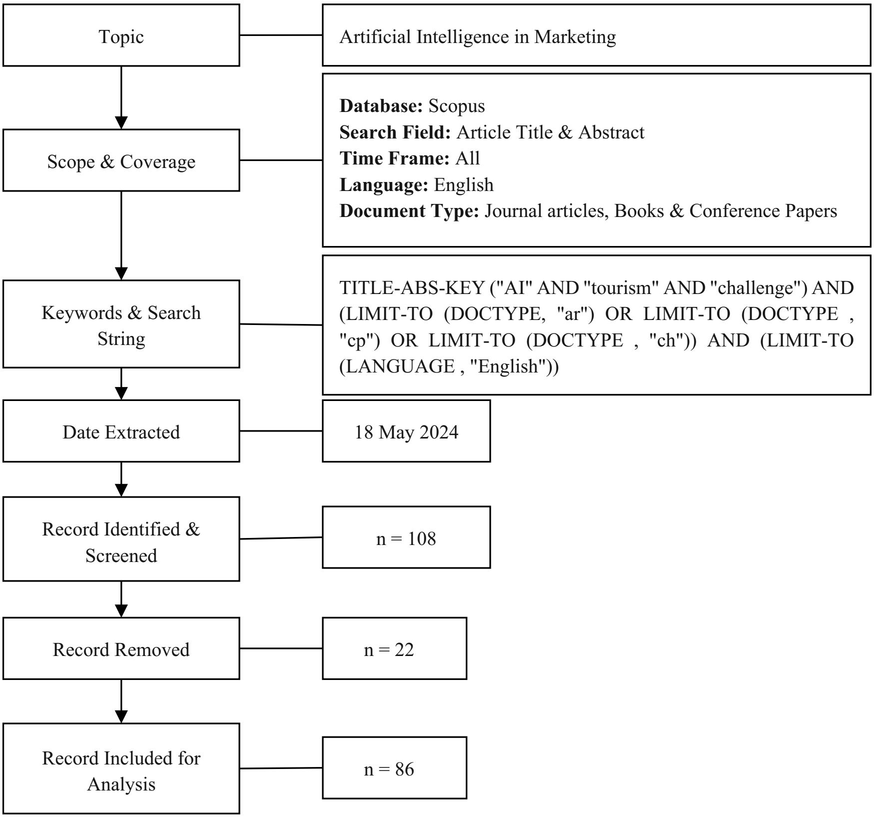


Fig. 1 Searching strategy flow chart. *Source* from: [39]

Journal articles, books, and conference papers were selected in order to emphasise academically rigorous content. In addition, this study targeted publications exclusively published in English in the purpose of or convenience and comprehensibility, therefore resulting in the remaining count of 108 papers.

After an extensive screening of the retrieved documents, duplicates or items irrelevant to the current subject have been eliminated. An in-depth review of titles and abstracts was conducted as screening technique, excluding papers not relevant to the subject domain. Following a manual data purification procedure in Office Excel 2019, a curated series of 86 papers was selected for additional literature review research.

4 Findings

4.1 Opportunities of AI in Tourism

AI Applications for Service Providers

Entities within the tourist business that offer products or services contains hotels, resorts, travel agencies, tour agencies, airlines, cruise and destination marketing organisations. These providers can use AI technologies to optimise operational efficiency, customise customer experiences, and enhance service delivery.

Optimisation. The deeper insights and well-informed business decisions for businesses ultimately leads to improved customer satisfaction and business performance. Studies conducted by [38, 50] illustrate the capability of AI and machine learning to predict hotel occupancy rates and optimise the room prices. This can result in more accurate forecasts, and eventually increase the hotel profitability. Furthermore, AI-powered sentiment analysis can offer a powerful means to comprehensively comprehend consumers' feedback [8]. Through the examination of customer online reviews on platforms such as TripAdvisor, AI has the capability to recognise both favourable and unfavourable feedbacks in terms of hotel services. This can help enable the identification of areas that need improvement, the management of online presence, and the development of precise marketing strategies [8].

Personalisation. AI can utilise the digital footprints to create marketing content that deeply resonates with individuals on a personal level. The information provided comes from sociodemographic characteristics, encompassing factors such as age, gender, personal interests, and specific details about the purpose and destination of the trip [16]. Crucially, AI has the ability to utilise current data from various places to offer accurate suggestions for travel routes, dining options, and lodging. This greatly enhances the overall experience of travellers [51, 55]. AI can significantly enhance the efficiency of marketing efforts and greatly improve the satisfaction and engagement of travellers through its advanced capabilities.

Automation. AI technologies, such as chatbots and service robots, are transforming customer service and operational efficiency in the hospitality industry. Chatbots can offer 24/7 help, dealing with various missions from common inquires to easy bookings. This boosts service speed, cuts down response times, and makes guests more satisfied and more loyal. Unlike the humanoid-looking robots employed at Henn-na Hotel, a recent study by [34] proposed a vision for the use of AI dogs in scenarios such as restaurants, city tour guides, extended-stay resorts, and trade shows.

AI Applications for Service Consumers

Pre-Travel Stage. It has been found that 74% of people plan their trips online, and over 45% use smartphones for this [43, 46]. Before going on a trip, visitors will use AI tools like ChatGPT to obtain detailed information about historical sites, museums, and cultural attractions and even create customised plans based on their personal interests [26]. Having access to accurate and reliable information helps travellers make better decisions and improves their travel experience [28]. AI can decrease tour charges by eliminating the demand for local guides to plan routes, which is especially valuable in the current economic climate, in which clients tend to control spending, particularly overseas occasions [13]. Moreover, AI tools can offer recommendations which are less popular but equally fascinating destination, helping to distribute tourist traffic more evenly and support sustainable tourism practices.

During The Trip. AI technology is vital for enhancing the travelling experience in multiple ways. AI programs like ChatGPT have can access to a huge amount of information, enabling them to provide travellers with up-to-date information such as news, weather, public transportation schedules, and any other details [56]. Moreover, Google Maps has become an indispensable tool for travellers with the help of AI and GPS technology. It provides real-time navigation and live updates about the traffic condition [46]. Utilising the features of live view, tourists can easily find the nearby shops, hotels, restaurants, and transportation junction, effectively acting as a local guide [46]. In addition, AI-powered language translation programs benefit travellers in foreign countries. These apps can translate signs, menus, and different textual content in real-time using the phone's camera, which can assist travellers interact with locals more easily [46]. This technology removes boundaries of language, thus reducing costs and improving the travel experience. Furthermore, AI is changing visitor experiences through innovative room technologies inside hotels and accommodation. Hotels employs technologies such as Amazon Alexa, allowing visitors to control lighting, temperature, and audio-visual systems and other aspects of the room environment—simply via voice commands [6, 16]. This enhances guest comfort and personalises the stay experience to a great extent.

4.2 Challenges of AI in Tourism

Technical Challenges

Risk of Technological Reliability. The reliability of AI and robotics is a significant concern which affects consumer adoption and trust. Consumer often has hesitation to use new technology because of worries about breaking down, being too complicated, or having other issues [52]. Breakdowns in robotic systems are fairly common. For example, a study [2] found that there were about 8,061 incidents involving problems with robots used in surgeries in the U.S. over the past ten years (as cited in [52]). In the hospitality industry, the fear that service robots might malfunction, such as delivery robots possibly spilling food, can greatly reduce people's willingness to use them [52]. Also, chatbots often can only handle basic questions, and customers need human help for more complicated issues [33], as cited in [46]. Many AI systems are viewed as "black boxes" because their inner workings are not clear, which weakens users' trust and understanding [3].

Data Availability and Updates

AI models need large and varied datasets to make accurate predictions [3]. Data is becoming increasingly important for gaining a competitive edge, as companies are creating advanced algorithms to use this data in new ways [19]. However, getting high-quality and representative data can be difficult, especially in areas with limited information or strict privacy rules. To make AI systems more accurate and reliable, problems with data collection, labelling, and organisation need to be solved [3]. Currently, only huge technology companies, including Google, Facebook, Amazon, and Booking.com, are capable of generating datasets of this magnitude [8]. The high costs associated with big data are a significant challenge for numerous organisations [47]. Furthermore, the dynamic character of tourism data makes it difficult for businesses to handle the enormous volumes of data required for AI systems such as ChatGPT to offer real-time information fast [20]. AI systems must learn and use data by constant data updates, which calls for significant time and money [5]. As a result, first ensuring the best performance of AI in tourism is effective information management and updating strategies.

Financial Barriers

Initial Deployment Cost. AI system development and first implementation call for large financial investment. Using Chatbot, for instance, entails building and maintaining hardware and software by means of skilled personnel, high-performance computing resources, and massive data centres, thus addressing the establishment and management of hardware and software [20]. Small and medium-sized businesses (SMEs) may find these initial investments required difficult. Although AI technology has great advanced capabilities, the high initial expenditure may hinder the adoption for small service providers [40]. Furthermore, complex questions still call for human involvement even if chatbots and robots can partially replace human labour [33]. For those companies, this could complicate their cost benefit study.

Maintenance And Operational Cost. Except the initial deployment cost, operation and maintenance of AI systems also calls continuous investments. Maintaining high-performance computing resources, managing data centres, and ensuring the AI models are current and operating as expected constitute part of these ongoing investments [8, 20]. Given the cost of keeping such sophisticated systems, SMEs could find these continuous costs taxing. For instance, low customer acceptance and high running costs drove many of the "robot hotels" in Japan to have to retire their robots [7]. Furthermore, including AI initiatives into internal corporate operations comes with additional expenses ranging from subscription fees for premium versions to the cost of recruiting programmers to train and apply language models [20]. These ongoing expenses draw attention to the financial difficulties companies have maintaining AI technologies.

Legal Considerations

Privacy Concerns. The use of tourism business generates serious privacy issues, mostly related to the collecting and handling of delicate personal and travel data.

Growing consumer privacy concerns [22] can hinder the general deployment of these technologies. Chatbots and AI-pushed systems in hotels often save past purchase histories and travel information, so raising questions about data privacy and security [29, 46]. Apart from that, AI systems such as ChatGPT, which handle sensitive data, have great hacking and data leaks security concerns [30]. This can damage clients and erode their confidence in these systems. These privacy issues demand strong cybersecurity systems and moral data management practices to guard user data.

Regulatory Compliance. Adopting AI in the travel sector calls for rigorously following regulatory compliance. Following the legal frameworks controlling data protection and privacy helps to avoid legal consequences. Particularly in sensitive areas like finance and military applications, regulatory bodies could pressure rigorous policies on collecting, storing, and using personal information [18, 26, 46]. When companies try to include AI technologies into their activities, they expose legal hazards because of these uncertain legal terrains. Therefore, following these rules is absolutely necessary to stay out of legal hot land.

Socio-Ethical Dilemmas

Human Interaction. Although some customers are optimistic about AI, numerous customers express concerns about losing the "human touch," which may bring about outrage, confusion, or disappointment [10]. Customers who have established strong relationships with their product or service providers may even perceive insulted when AI is applied instead of human interaction. Research suggests that clients presently prefer receiving services from human staff rather than robots, although the human service is inferior and more expensive [21]. This is noteworthy for those tourism and hospitality businesses which typically emphasises human connection and relationship establishment. Decreased customer satisfaction, loyalty, and overall engagement may arise due to failure to address these concerns, which ultimately impacting the industry's reputation and revenue.

Job Displacement. The rise of AI technologies like ChatGPT can significantly influence automating routine and repetitive jobs, which could result in mass unemployment and reskilling requirements [1]. For example, traditional roles such as travel consultants, customer service agents, concierge agents, and guides may be replaced by AI technologies [9]. Highly populated countries like China, India, and Brazil might suffer from this influence most, as a significant portion of the population is employed in these industries [1]. As AI continues to advance, it will be crucial to deal with the challenges of job displacement and ensure that workers have opportunities for reskilling and adjustment to new jobs within the industry.

5 Future Research

5.1 *Understanding Tourist Acceptance of AI.*

Acceptance among tourists towards AI is an essential field of research that may provide valuable insights into the integration of AI in the tourism industry. Research opportunities involve investigating the various factors that can influence users' acceptance of AI, including Chatbot-specific attributes (language fluency, level of personalisation, accuracy, etc.), user-specific factors (innovativeness, sociodemographic factors, risk-taking behaviour, etc.), and societal factors (social acceptance, skill development, new job creation, etc.), offer forward-thinking perspective [20]. Specific research questions could include:

RQ1. How do different Chatbot-specific attributes influence tourist acceptance of AI in hospitality?

RQ2. What role do sociodemographic factors play in tourists' willingness to use AI services?

RQ3. How does societal acceptance of AI impact its adoption in the tourism sector?

Furthermore, examining the impact of user experiences with AI on the willingness of tourists to adopt AI could involve investigating the potential of AI to be utilised to enhance the tourist experience. For instance:

RQ4. Do visitors with prior experience with Chatbots exhibit a higher propensity to persist in their usage?

RQ5. What are the key factors that influence the continued use of Chatbots in tourism?

5.2 *Enhancing Tourism Experiences with AI*

Exploring how AI can enhance the overall travel experience is another vital area of research. Understanding the benefits of AI in creating and delivering experiences, and investigating how AI can be employed to improve these experiences, will be valuable for both theoretical insights and practical application [20]. Specific research questions could include:

RQ6. What are the key advantages of AI in enhancing tourist experiences?

RQ7. How can AI be used to create more immersive and interactive tourist experiences?

By tracking and analysing tourists' interactions with AI, researchers and professionals can identify the factors that improve the overall tourism experience.

RQ8. How do tourists' expectations of AI differ across various segments?

RQ9. Does AI fulfil the expectations of various tourist categories, and what factors influence its use in these segments?

5.3 Role of AI in Travel Decision-Making Processes

A thorough empirical investigation is required to examine AI's potential impact on travellers' decision-making processes. Studies like these can look into how AI changes the way travellers search for and handle information by providing them with a wider variety of personalised travel details, even about less well-known places [20]. Some possible research questions might be:

RQ10. How does AI affect the way travellers search for information and make decisions?

RQ11. What is the interaction between AI and other sources of travel information, and how do they collectively impact travellers' decisions?

5.4 Enhancing Data Encryption and Security Protocols

Future research should prioritise the development and implementation of advanced encryption techniques and robust security protocols to protect sensitive data from unauthorised access, hacking, and leaks. Specific research questions could include:

RQ12. What are the most effective encryption techniques for securing data in the tourism industry?

RQ13. How can AI models incorporate privacy-preserving mechanisms such as differential privacy and federated learning?

Research should also assess how well these methods work in stopping data breaches in real-life situations within the tourism industry.

5.5 Evaluating Regulatory Compliance and Legal Frameworks

Balancing the efficiency gains of AI with safeguarding job prospects for human workers presents a difficult problem [27]. Specific study questions might be:

RQ16. What policies might help to reduce employment losses in the tourism industry brought on by AI?

RQ17. How might programs for workforce retraining support those impacted by AI adoption?

Deeper knowledge of sustainable solutions for this sector will come from analysing the long-term socioeconomic consequences of AI-driven job displacement in different worldwide environments.

5.6 Addressing Job Displacement and Reskilling Considerations

The use of AI in tourism brings up important concerns about job loss. There is a complex challenge between balancing the efficiency benefits of AI with preserving job positions for human workers [27]. Specific research questions could include:

RQ16. What strategies can mitigate job losses due to application of AI in the tourism sector?

RQ17. How can workforce retraining programs be designed to support employees affected by AI adoption?

Analysing the long-term socioeconomic effects of AI-driven job displacement in distinct international settings will provide deeper understanding into sustainable solutions for the industry.

6 Conclusion

Unquestionably, artificial intelligence (AI) is important for the tourism business since it quickly changes it by improving efficiency, personalisation, and general quality of travel experiences. From customer service to resource management, AI technologies including machine learning, robotization, deep learning and big data have great power to transform many facets of travel. Using these technologies enables travel companies to keep a competitive advantage in a changing market, simplify processes, and offer more customised services.

However, there are also a lot of obstacles in the way of AI's adoption in the tourism industry. It is necessary to address technical concerns about the dependability of AI systems, the accessibility of high-quality data, and the difficulty of managing AI-driven processes. The high upfront and ongoing maintenance costs associated with these technologies may discourage small and medium-sized businesses (SMEs) from implementing them. Further challenges that need to be carefully navigated are legal and socio-ethical ones, such as data privacy, regulatory compliance, and the possibility of job displacement.

Future studies should focus on figuring out how tourists feel about AI, how AI can enhance travel experiences, and how AI influences travellers' decision-making. Furthermore, it is critical to improve data encryption and security procedures, assess legal and regulatory compliance, and deal with issues related to reskilling and job displacement. These studies will provide insightful information and practical tactics to shape an ethical and inclusive future for AI in tourism, ensuring that technological breakthroughs benefit the sector and its workforce.

Through investigating these research paths, scientists and professionals will be able to fully grasp the revolutionary possibilities of AI in the travel industry, so guiding the acceptance of more creative and competitive industrial methods. Integrating AI will become increasingly important as the travel and hospitality sector develops in improving the quality of travel experiences and guaranteeing sectoral sustainable growth.

There is great long-term impact of AI in travel. AI can redefine operational and strategic frameworks of tourism companies as well as change consumer interactions and service delivery. Policymakers, industry professionals, and researchers are absolutely essential to fully use the possibilities of AI by means of cooperative efforts. This forward-looking approach emphasises the need of addressing the listed issues and opens the path for AI-driven innovations that might result in a more dynamic, effective, and strong travel sector.

Finally, this paper emphasises the importance of AI in the hotel, travel, and tourist sectors as well as its practical uses. Emphasising the need of thorough research and cooperative efforts will help AI technologies be used in ways that maximise their advantages and minimise related risks, so benefiting suppliers as well as consumers.

References

1. Agrawal, A., Gans, J., Goldfarb, A.: ChatGPT and How AI Disrupts Industries, https://hbr.org/2022/12/chatgpt-and-how-ai-disrupts-industries, last accessed 2024/08/01. (2022).
2. Alemzadeh, H., Raman, J., Leveson, N., Kalbarczyk, Z., Iyer, R.K.: Adverse events in Robotic surgery: A retrospective study of 14 years of FDA data. PLOS ONE. 11. (2016).
3. Aliyah, Lukita, C., Pangilinan, G.A., Chakim, M.H.R., Saputra, D.B.: Examining the Impact of Artificial Intelligence and Internet of Things on Smart Tourism Destinations: A Comprehensive Study. Aptisi Transactions on Technopreneurship (ATT). 5, 135–145 (2023).
4. Almasi, S., Cosmas, A., Cowan, S., Ellencweig, B.: The promise of travel in the age of AI, https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/the-promise-of-travel-in-the-age-of-ai, last accessed 2024/07/30.
5. Arif, T.B., Munaf, U., Ul-Haque, I.: The future of medical education and research: Is ChatGPT a blessing or blight in disguise? Medical Education Online. 28, (2023).
6. Basiri, A., Amirian, P., Winstanley, A., Moore, T.: Making tourist guidance systems more intelligent, adaptive and personalised using crowd sourced movement data. Journal of Ambient Intelligence and Humanized Computing. 9, 413–427 (2017).
7. Buhalis, D.: Technology in tourism-from information communication technologies to eTourism and smart tourism towards ambient intelligence tourism: a perspective article. Tourism Review. 75, (2020).
8. Bulchand-Gidumal, J., Secin, E.W., O'Connor, P., Buhalis, D.: Artificial intelligence's impact on hospitality and tourism marketing: exploring key themes and addressing challenges. Current Issues in Tourism. 1–18 (2023).
9. Carvalho, I., Ivanov, S.: ChatGPT for tourism: applications, Benefits and Risks. Tourism Review. 79, (2023).
10. Castillo, D., Canhoto, A.I., Said, E.: The dark side of AI-powered service interactions: Exploring the process of co-destruction from the customer perspective. The Service Industries Journal. 41, 900–925 (2020).
11. De Mauro, A., Greco, M., Grimaldi, M.: What is big data? A consensual definition and a review of key research topics. What Is Big data? a Consensual Definition and a Review of Key Research Topics. 1644, (2015).
12. Deloitte: 2024 travel industry outlook. (2024).
13. Deloitte: Travel Week Insight Report 2024. (2024).

14. Essien, A., Chukwukelu, G.: Deep learning in hospitality and tourism: a research framework agenda for future research. International Journal of Contemporary Hospitality Management. 34, 4480–4515 (2022).
15. Future Data Stats: Artificial Intelligence in Hospitality and Tourism Market Size, Share, Trends & Competitive Analysis By Type (Natural Language Processing Applications, Machine Learning Algorithms, Computer Vision and Image Recognition, Chatbots and Virtual Assistants, Recommendation Systems, Sentiment Analysis) By Application; By End-User; By Regions, and Industry Forecast, Global Report 2023–2030. (2023).
16. Gajdošík, T., Marciš, M.: Artificial Intelligence Tools for Smart Tourism Development. Advances in Intelligent Systems and Computing. 392–402 (2019).
17. Gandomi, A., Haider, M.: Beyond the Hype: Big Data Concepts, Methods, and Analytics. International Journal of Information Management. 35, 137–144 (2015).
18. Government of Canada: Privacy Act Modernization: A Discussion Paper - 1. Privacy principles and modernized rules for a digital age, https://www.justice.gc.ca/eng/csj-sjc/pa-lprp/dp-dd/modern_1.html, last accessed 2024/08/02.
19. Grossberg, K.A.: The new marketing solutions that will drive strategy implementation. Strategy & Leadership. 44, 20–26 (2016).
20. Gursoy, D., Li, Y., Song, H.: ChatGPT and the hospitality and tourism industry: an overview of current trends and future research directions. Journal of Hospitality Marketing & Management. 32, 1–14 (2023).
21. Gursoy, D., Cai, R.: Artificial intelligence: an overview of research trends and future directions. International journal of contemporary hospitality management. (2024).
22. Hu, Y., Min, H. (Kelly): The dark side of artificial intelligence in service: The "watching-eye" effect and privacy concerns. International Journal of Hospitality Management. 110, 103437 (2023).
23. IBM: What is Deep Learning?, https://www.ibm.com/topics/deep-learning.
24. Ivanov, S., Webster, C.: Robots in tourism: A research agenda for tourism economics. Tourism Economics. 26, 135481661987958 (2019).
25. Ivanov, S., Webster, C.: Willingness-to-pay for robot-delivered tourism and hospitality services – an exploratory study. International Journal of Contemporary Hospitality Management. 33, 3926–3955 (2021).
26. Jia, S., Chi, O.H., Martinez, S.D., Lu, L.: When "Old" Meets "New": Unlocking the Future of Innovative Technology Implementation in Heritage Tourism. Journal of Hospitality & Tourism Research. (2023).
27. Joseph, A.: Artificial Intelligence in #Travel — Top 7 Challenges, https://medium.com/publishfive/artificial-intelligence-in-travel-top-7-challenges-f860702acb05, last accessed 2024/07/11.
28. Kadyan, V., Singh, T.P., Ugwu, C.: Deep Learning Technologies for the Sustainable Development Goals. Springer Nature, Singapore (2023).
29. Kannan, P., Bernoff, J.: The future of customer service is AI-human collaboration, https://sloanreview.mit.edu/article/thefuture-of-customer-service-is-ai-human-collaboration/, last accessed 2014/07/10. (2019).
30. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., Stadler, M.: ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences. 103, (2023).
31. Kim, H., Kam, K., Shin, S., Li, J.: Artificial Intelligence in Hospitality and Tourism: Insights From Industry Practices, Research Literature, and Expert Opinions. Journal of Hospitality & Tourism Research. (2024).
32. Le, T.H., Arcodia, C., Novais, M.A., Kralj, A.: Proposing a systematic approach for integrating traditional research methods into machine learning in text analytics in tourism and hospitality. Current Issues in Tourism. 24, 1640–1655 (2020).

33. Lommatzsch, A.: A Next Generation Chatbot-Framework for the Public Administration. In: Hodoň, M., Eichler, G., Erfurth, C., and Fahrnberger, G. (eds.) Communications in Computer and Information Science. pp. 127–141. Springer (2018).
34. Lu, Y. (Darcy), Liang, Y., Wang, Y.-C.: AI dogs vs. real dogs and human-like robots: clarification, conceptualization, and applications in tourism and hospitality settings. Journal of Hospitality and Tourism Technology. (2024).
35. Mariani, M., Baggio, R., Fuchs, M., Höepken, W.: Business intelligence and big data in hospitality and tourism: a systematic literature review. International Journal of Contemporary Hospitality Management. 30, 3514–3554 (2018).
36. Mariani, M., Baggio, R.: Big data and analytics in hospitality and tourism: a systematic literature review. International Journal of Contemporary Hospitality Management. ahead-of-print, (2021).
37. Maulana, N.: Research Trends in Marketing Science Before COVID-19 Outbreak: A Literature Review. Management & Marketing. Challenges for the Knowledge Society. 15, 514–533 (2020).
38. Millauer, T., Vellekoop, M.: Artificial intelligence in today's hotel revenue management: opportunities and risks. Research in Hospitality Management. 9, 121–124 (2019).
39. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G.: Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. PLoS Medicine. 6, (2009).
40. Murphy, J., Hofacker, C., Gretzel, U.: Dawning of the Age of Robots in Hospitality and Tourism: Challenges for Teaching and Research. European Journal of Tourism Research. 15, 104–111 (2017).
41. Norfleet, N.: Robotic software sweeping large accounting firms and clients, https://www.virginislandsdailynews.com/robotic-software-sweeping-large-accounting-firms-and-clients/article_3af86f4f-f09e-5a51-a340-3ff9befc48af.html, last accessed 2024/07/03.
42. Núñez, J.C.S., Gómez-Pulido, J.A., Ramírez, R.R.: Machine learning applied to tourism: A systematic review. Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery. (2024).
43. Peranzo, P.: AI assistant: The Future of Travel Industry with the Increase of Artificial Intelligence, https://imaginovation.net/blog/the-future-of-travel-with-the-increase-of-ai/, last accessed 2024/07/03. (2019).
44. PwC: Robotization, https://www.pwc.com/sk/en/dictionary/robotization.html, last accessed 2024/06/22.
45. Ranjbari, M., Shams Esfandabadi, Z., Scagnelli, S.D.: A big data approach to map the service quality of short-stay accommodation sharing. International Journal of Contemporary Hospitality Management. 32, 2575–2592 (2020).
46. Samala, N., Katkam, B.S., Bellamkonda, R.S., Rodriguez, R.V.: Impact of AI and robotics in the tourism sector: a critical insight. Journal of Tourism Futures. ahead-of-print, (2020).
47. Sivarajah, U., Kamal, M.M., Irani, Z., Weerakkody, V.: Critical analysis of Big Data challenges and analytical methods. Journal of Business Research. 70, 263–286 (2017).
48. Tribe, J.: The Economics of Recreation, Leisure and Tourism. Routledge (2020).
49. Veiga, C., Santos, M.C., Águas, P., Santos, J.A.C.: Are millennials transforming global tourism? Challenges for destinations and companies. Worldwide Hospitality and Tourism Themes. 9, 603–616 (2017).
50. Viverit, L., Heo, C.Y., Pereira, L.N., Tiana, G.: Application of machine learning to cluster hotel booking curves for hotel demand forecasting. International Journal of Hospitality Management. 111, 103455–103455 (2023).
51. Wang, W., Kumar, N., Chen, J., Gong, Z., Kong, X., Wei, W., Gao, H.: Realizing the Potential of Internet of Things for Smart Tourism with 5G and AI. IEEE Network. 34, 295–301 (2020).
52. Wang, X., Zhang, Z., Huang, D., Li, Z.: Consumer resistance to service robots at the hotel front desk: A mixed-methods research. Tourism Management Perspectives. 46, 101074 (2023).
53. World Travel & Tourism Council: Travel& Tourism Economic Impact 2022. (2022).
54. Wu, D.C., Song, H., Shen, S.: New developments in tourism and hotel demand modeling and forecasting. International Journal of Contemporary Hospitality Management. 29, 507–529 (2017).

55. Yang, X., Zhang, L., Zhou, F.: Personalized Tourism Recommendations and the E-Tourism User Experience. Journal of Travel Research. 63, (2024).
56. Zhou, P.: Unleashing ChatGPT on the Metaverse: Savior or Destroyer? (2023).

Artificial Intelligence (AI)-Enabled Mobile Fitness Apps and Goal Attainment: Systematic Literature Review

Zhang Yanyan, Noorminshah A.Iahad, and Ahmad Fadhil Yusof

Abstract With the widespread adoption of mobile devices and advancements in artificial intelligence (AI), AI-driven mobile fitness applications (MFAs) have become a part of everyday life for many individuals. Existing research has not thoroughly examined how AI-powered mobile fitness applications (MFAs) influence users' achievement of their goals. We conducted a systematic review to assess the effects of AI-powered mobile fitness applications (MFAs) on achieving fitness goals. Our methodology included a standardized search across four leading scientific databases, with filters applied for language and type of article. This research focused on the methods, research questions, and applications. Drawing on the systematic literature review (SLR), this paper provides a comprehensive demonstration of the application of AI to different goal types and outcomes. Challenges and opportunities are also identified in this research.

Keywords Artificial intelligence · Mobile fitness apps · Systematic literature review · Goal attainment

1 Introduction

Artificial Intelligence (AI) plays a significant role in the context of health and fitness. In the health and fitness sector, technological innovations such as artificial intelligence and the Internet of Things have reshaped the way users pursue their goals in fitness, from goal setting to goal achievement. Personalized services are

Z. Yanyan (✉)
Faculty of Management, Universiti Teknologi Malaysia, 81310 Johor, Malaysia
e-mail: yanyan@ieee.org

N. A.Iahad
Department of Information Systems, Faculty of Management, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

A. F. Yusof
Faculty of Computing, Universiti Teknologi Malaysia, 81310 Johor, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_10

also constantly improving through enhanced instant feedback and advanced activities. Current fitness software provides personalized exercise plans based on physical conditions and suggests diets using image recognition. Artificial Intelligence-enhanced fitness software enhances user experience and increases participation, which in turn promotes health improvement. However, for these benefits to truly take effect, AI implementation challenges must be resolved and ensure their effective utilization.

This paper investigates how AI mobile fitness apps can assist individuals in meeting fitness goals. An in-depth analysis is performed of existing articles regarding AI applications, goal achievement strategies and any challenges. This paper categorizes fitness goals into three different types: psychological, physiological and interpersonal. This classification helps researchers better understand how AI in fitness software can support different goals. Personalization is one of AI's main strengths, which can encourage more people to use fitness software sustainably by offering tailored guidance and improving user experience. However, using AI in fitness software poses ethical concerns such as protecting data privacy or creating overly tailored recommendations.

This paper is organized as follows. Section 2 contains a literature review that introduces key aspects of Artificial Intelligence (AI), its subsets, relationships, and trends as well as features found within mobile fitness applications that utilize AI as well as goal attainment within fitness software applications. Research Methodology details the search strategy employed for selecting studies. It includes information regarding databases and keywords used as well as criteria to select or exclude studies from consideration. Section 4 discusses the findings in terms of using AI for mobile fitness apps, its effect on reaching three fitness goals, and any potential technical or ethical hurdles it might present. Finally, two additional sections address the limitations of this study and offer recommendations for future research before concluding the paper.

2 Literature Review

This section presents an analysis of the academic literature on the role of AI-enabled mobile fitness apps in achieving fitness goals. There are three sections. The first section provides an overview of Artificial Intelligence (AI) and explores various AI technologies, including machine learning, deep learning, and big data analytics. The second part summarizes the key features and functionalities of mobile fitness apps. And integration of AI in Mobile Fitness Apps (MFAs), highlighting its applications associated with implementing AI technologies to enhance user goal attainment. The third section provides a detailed analysis of the components of goals and the mechanisms of goal attainment within the context of mobile fitness apps.

2.1 Artificial Intelligence and Subsets

Rapid technological advancements have led to a new era of AI-enabled mobile applications, which are now being used by people of all backgrounds around the world. AI is increasingly being integrated into everyday life, becoming a critical component of modern living [3]. In recent decades, there have been many efforts to define artificial intelligence. Despite these attempts, there is still a lack of consensus among experts in various fields that contribute to artificial intelligence, such as mathematics and computer science. As a result, a universally accepted definition of AI has yet to be established [1].

Despite differing perspectives, the basic concept of AI has always been the creation of machines capable of human-like reasoning [25]. 2021, Sack explains AI and its related subsets. Machine learning (ML) is a subfield of artificial intelligence (AI) that aims to simulate experiential aspects of human intelligence, specifically the ability to learn, using computational algorithms [15]. Although views vary, the basic concept of artificial intelligence has always been to create machines that can reason like humans [25]. 2021, Sack explains AI and its related subsets. Machine Learning (ML) is a subfield of Artificial Intelligence (AI) that aims to simulate the empirical aspects of human intelligence, particularly the ability to learn, using computational algorithms [15]. The utilisation of big data helps train algorithms so that predictable patterns can be identified. Machine learning (ML) algorithms fall into two main categories: trained and untrained algorithms. Data contains features or "labels" that are incorporated into machine learning algorithms to facilitate the learning process. Non-cognitive machine learning methods should independently detect associations between inputs and do not require labels. A complex subfield of machine learning (ML) known as deep learning (DL) is often associated with neural networks. These networks use hierarchical structures to control inputs and outputs [15]. As inputs travel through multiple secretory layers, the brain's neurons make complex connections. Real-world applications include enterprise management in commercial banks [22] and semantic segmentation to extract information from images and videos [35].

In summary, deep learning technology has the potential to significantly change today's landscape. It serves as a robust computational engine that drives technological advancements, promotes automation, and supports the development of smart and intelligent systems. The objective of Industry 4.0 is closely aligned with the principles of deep learning, which facilitates the development of a highly automated and interconnected industrial environment [34].

2.2 Ai-Enabled Mobile Fitness App

GSMA [14] reports that mobile internet usage is on the rise, with 57% of the world's population, or 4.6 billion people, currently connected. Between 2017 and

2021, mobile users engaged in an increasing number of activities, particularly non-communications-related activities, during the pandemic lockdown. These activities included accessing health and government services, educational support, and applying for jobs. The global COVID-19 pandemic in 2020 has led to widespread shortages of medical and healthcare resources in almost every country. Consequently, there is an increasing recognition of the importance of maintaining a healthy lifestyle and incorporating regular physical activity into daily routines [7, 23].

mHealth applications are software integrated into smartphones that have the objective of enhancing health outcomes, research, and services. In contrast to health interventions delivered via laptops and computers, mHealth apps provide a more cost-effective and thus more accessible platform [28, 36]. Pires et al. [30] classifies mHealth apps into six categories: health record apps, lifestyle apps, remote counseling and monitoring apps, health education apps, apps for contacting healthcare providers, and diagnostic and treatment apps. Mobile fitness applications fall within the realm of lifestyle and health applications. Fitness refers to both overall health and vitality and maintaining an ideal body appearance through physical exercise [12]. Since the 1980s, "health-related physical fitness" (HRPF) has received considerable research interest—its components being the ability to perform physical activities, energy levels needed to complete daily tasks effectively, reduced disease risks due to sedentary living [10].

Fitness apps have been one of the most popular mHealth apps, providing users with access to tools to monitor their eating habits and participate in physical activities that promote a healthier lifestyle [7]. Furthermore, mobile health apps may recommend various fitness programs tailored specifically to users' needs such as strength training or Aerobic workouts [7]. This help users to choose from various suitable choices [40].

The demand for MFPs has increased significantly as individuals integrate epidemic prevention measures with their personal fitness goals. It is critical to consider the economic impact of MFPs and their potential benefits to individuals and society [20]. Multifunctional mobile apps are progressively dominating the business mobile apps market, growing rapidly in the field and becoming an important focus of academic research [7]. Artificial Intelligence (AI) algorithms are currently being incorporated into MFA to improve user experience and engagement with fitness services. Unlike traditional MFA, AI-enabled MFA strives to provide smarter and more personalised fitness services to users engaged in physical activity and exercise [13].

2.3 Users' Fitness Goal

Goals play a key role in providing direction for individuals to achieve various goals such as weight loss, diet control, and stress reduction. Once these goals have been achieved, individuals often terminate the utilisation of fitness techniques, indicating that the utilisation of fitness technology is inherently goal-directed. The literature on goal-directed behaviour defines this type of behaviour as actions oriented towards achieving specific outcomes. This process may be conceived as a sequence of stages, commencing with the establishment of a goal and subsequent formulation of the goal intention, followed by the implementation of a plan of action and ultimately the attainment of the desired outcome [5, 6]. Despite the prevalence of mobile fitness apps, the influence of AI-powered MFAs utilisation on the process of fitness goal attainment remains unexamined, many studies focus on goal-related components. For example, [20] demonstrated that the intelligence and anthropomorphisation of mobile fitness apps (MFAs) significantly influence users' willingness to continue utilising these apps. This effect is mediated by two mediating factors: goal difficulty and specificity. This study corroborates the significance of well-defined and challenging goals from a goal-setting perspective.

Additionally, fitness goal attainment is a direct result of engaging in fitness activities and an indirect result of using fitness technologies. Therefore, examining the relationships between fitness technology use, participation in fitness activities, and fitness goal attainment can provide valuable insights that enhance the positive effects of fitness technology on goal attainment. While it is reasonable to assume that individuals use fitness technologies to achieve their fitness goals, the specific impact of AI features in mobile fitness apps on goal attainment is not yet well understood. To optimise their investment in corporate wellness programmes, organisations need to gain a deeper understanding of how individuals interact with fitness technologies and achieve their fitness goals [21].

3 Methodology

This research aims to investigate the following questions: (1) What are the primary applications of AI technologies in MFAs? (2) What are the consequences of using AI-enabled MFAs on user goal attainment? (3) What are the significant challenges faced in the integration of AI technologies in MFAs?

3.1 Search Strategy

The search terms for AI-enabled MFAs and the search strategies for various databases are detailed in Table 1. To identify the most pertinent articles in the field of machine learning, big data, and artificial intelligence and their applications in the context of mobile fitness apps, the following keywords should be employed in the search: machine learning; big data; artificial intelligence, and; mobile fitness app; goal attainment. In encompassing this interdisciplinary study, which covers a broad range of disciplines including marketing, health, and psychology, a variety of databases were selected to facilitate the retrieval of relevant information. These included Web of Science, Scopus, ScienceDirect, and Emerald. The search was conducted on 18 May 2024.

Table 1 Search strategy

Category	Search term
AI	(“data mining” OR “artificial intelligence” OR “big data” OR “machine learning”)
MFA	(“Mobile Fitness App” OR “smartphone” OR “Mobile phone”)
Goal attainment	(“goal attainment” OR “goal achievement” OR “meeting goals”)
Combine	1 2 and 3

3.2 Selection and Exclusion Criteria

In conducting this review, our focus was on research studies published in peer-reviewed journals, with dissertations and abstracts intentionally excluded. We did not consider grey literature, including evaluation reports, annual reports, articles in non-peer-reviewed journals, and other similar publications. The inclusion criteria for the articles were: (i) research articles, review articles, conference papers, and book chapters; (ii) publications in English; (iii) studies involving any type of AI application in the context of fitness apps; and (iv) those measuring specific AI applications during goal pursuit. The exclusion criteria were: (i) congress proceedings, books, or other types of publications; (ii) studies that did not use AI-enabled mobile fitness apps; and (iii) unavailability of the full paper.

3.3 Data Extraction and Synthesis

Figure 1 shows the [27] flow diagram, which follows the PRISMA methodology commonly used in systematic reviews of this type. The initial database search yielded 460 results, and the reviewers conducted a comprehensive check for duplicates and titles, followed by an abstract review and full-text review based on the inclusion and exclusion criteria. The screening process involves evaluating selected articles based on defined criteria that include the following requirements: the writing language is English, AI-powered fitness application, involves user goal setting as well as goal completion. The reasons for exclusion mainly include language factors, irrelevant articles i.e. traditional fitness software, and the full text does not deal with goal completion. The data flow diagram is shown in Fig. 1. Eventually, 27 articles were selected for detailed analysis. These transcripts can provide information about how AI technology in mobile fitness apps affects users’ goal achievement. As a result of this selection process, the final analyses were based on high-quality research.

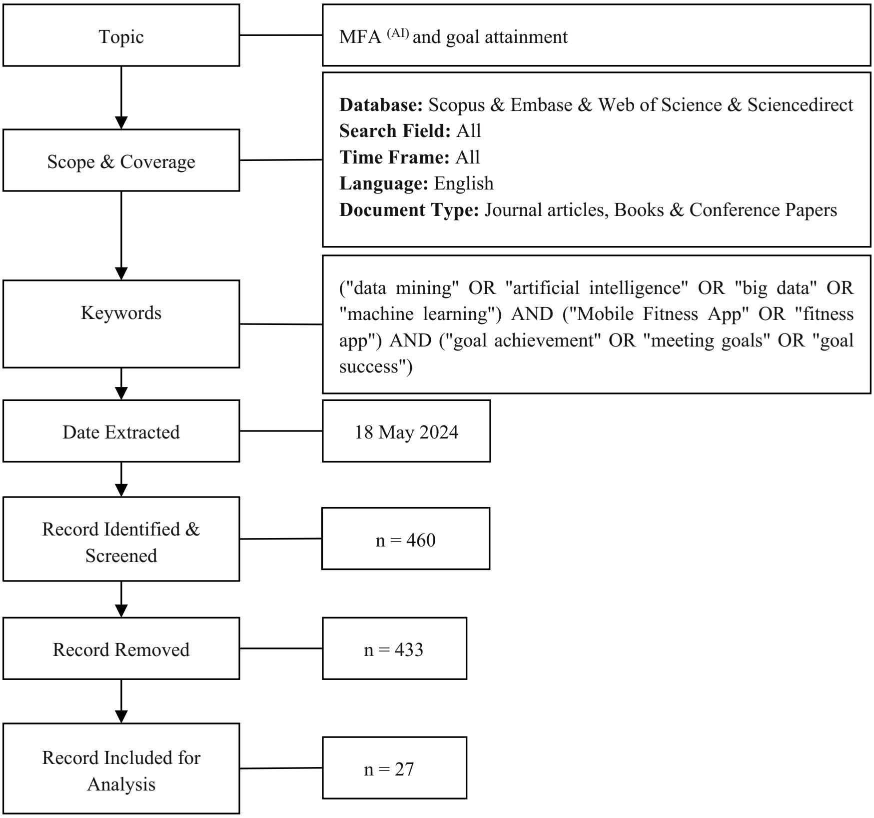


Fig. 1 Flow diagram. *Source* [27]

4 Results

This section analyses the main research results on AI mobile fitness applications. The first section mentions specific applications of AI in the context of mobile fitness apps. The second section explores how AI technology is applied to achieve these different goals. The third section shows the outcomes of AI fitness apps in achieving these goals using metrics from the perspective of continued use intention, habit formation, and satisfaction. The final section discusses the challenges of MFA (AI), including privacy issues, overload alerts, and cost.

4.1 AI in the Mobile Fitness App

As technology advances, the use of AI technology is becoming more widespread. the critical functionalities and the services and technologies associated with AI-enabled mobile fitness apps will be discussed in this section.

Critical functionalities. Each mobile fitness app consists of several components that determine its functionality. AI enhances these apps by enabling recognition (identifying types of activities or foods), monitoring (tracking user progress), personalisation and recommendations (offering tailored plans, guidelines, and suggestions based on user data and progress), similar to a professional coach [2]. The primary objective of many studies has been activity recognition, including robotic implementations. For example, Oyibo et al. highlighted that considerable research efforts have focused on this goal. Numerous AI-enabled mobile fitness apps (MFAs) used different combinations of behaviour change techniques (BCTs) [16]. Goal setting and self-monitoring were the two most frequently combined behavioral change tactics (BCTs). This combination demonstrates the app's effectiveness at setting specific fitness goals and tracking progress through features such as logging completed workouts, monitoring performance metrics, and tracking personal records [4]. Fitness apps are excellent tools for meeting individual users' needs and progress, which keeps people engaged, motivated, and committed to long-term changes [13]. Artificial Intelligence (AI) makes them even better by fine-tuning algorithms that adjust training intensity based on how each user is progressing, creating a coaching experience even more tailored and individualized for them. Integrating user preferences and progress into AI systems can provide more tailored recommendations and adjustments, encouraging user progress while building self-confidence. AI-enabled mobile fitness apps use objective and quantitative big data analyses to inform decision-making and help users create effective fitness regimens. These AI algorithms contain numerous functionalities, including training session detection, injury recovery support, training plan prediction as well as professional assistance services.

Services and Technologies. AI-powered mobile fitness apps use advanced services and technologies to enhance the user experience. These apps access both remote and local libraries. These devices feature advanced technologies like cameras, motion sensors (accelerometers and gyroscopes), GPS for location tracking and microphones; more advanced models may also include photos, vibrations, networking audio phone functions as well as storage functions. These features enable the collection of various forms of data, such as sensor data, images, voice messages, and text. Once collected, this data is then processed using machine learning algorithms to gain valuable insights and provide personalized fitness recommendations [2]. Digital technologies, including artificial intelligence-powered mobile fitness apps with natural language processing capabilities, enable users to engage in and monitor physical activity through wearable sensors and smartphones. These apps are intended to keep users informed about their progress, providing instantaneous feedback based on physical activity data [4]. AI-powered fitness applications leverage natural language processing to make their AI feel more lifelike by incorporating vision and hearing.

Furthermore, these apps offer timely feedback using real human voice prompts to correct and encourage user behavior, ultimately making AI appear more human-like [20]. Farrokhi et al. [13] summarizes intelligent fitness system which enables users to engage in fitness activities and store exercise data within their profiles. Artificial intelligence algorithms analyze this data, making recommendations based on results. Furthermore, users receive tailored advice based on their profile and performance that is stored for future reference.

4.2 MFA(AI) and Fitness Goal Attainment

Goal attainment is initiated by the pursuit of the desired goal. This concept encompasses the setting of goals, reflection, feedback, and the promotion of self-efficacy [37]. The significance of goal types may differ depending on the context in question. James et al. [18] present a framework for understanding the exercise goals that are relevant to the use of fitness technology. They define these goals and discuss whether they are commonly viewed as intrinsic or extrinsic. The authors categorise the goals into three main groups which are psychological goals, interpersonal goals, and physical goals. The following section will examine the utilisation of AI in the context of mobile fitness applications and goal attainment.

Psychological goals. As indicated by [18], there is a positive correlation between the psychological goals employed in fitness techniques and subjective vigor. Given that psychological goals have historically been regarded as intrinsic, this indicates that pursuing these goals can positively impact an individual's overall well-being and vitality. Psychological goals can be classified into three categories: enjoyment, challenge, and stress management. A multitude of studies have demonstrated the efficacy of providing emotional support to users. To illustrate, the incorporation of gamification design and elements into mobile fitness applications has been demonstrated to enhance user enjoyment. The incorporation of gaming elements into a user experience provides a diversion from the otherwise uninspiring or arduous objectives, thereby fostering the routine utilisation of motivational technology [29]. For example, game features such as virtual badges, prizes, and incentives, related to user's activities, lead to goal achievement [29, 39].

The relationship between enjoyment and life satisfaction has been demonstrated by [17], who designed an item to reflect this relationship. Regarding challenges, AI-enabled mobile fitness apps allow users to share their behaviors on social media or in forums. This can also involve features like a buddy system or leaderboard [39]. Stress management aims to provide space for reflection, reduce tension, manage stress, and release tension [24]. Greater well-being is strongly associated with both immediate and long-term health advantages, including stress reduction, lowered risk of disease and illness, enhanced immune function, and increased longevity [31]. A smaller number of studies have a direct focus on stress management. However, [17] examined stress from a different perspective, indicating that users may experience feelings of being overwhelmed by the abundance of social information, discussions, and comments, even when utilising fitness apps to adopt a healthier lifestyle. If social engagement through fitness apps is not effectively managed, it can result in stress and anxiety rather than providing relaxation and enjoyment.

Interpersonal goals. The interpersonal goals of social recognition, affiliation, and competition are focused on the development of interpersonal relationships [18]. The explanation provided by [17] can be further enhanced by the inclusion of the observation that socially oriented features inherent to fitness apps encompass elements that facilitate user interactions, including coaching, social comparisons, and networkability. Social recognition involves demonstrating accomplishments, comparing abilities, and gaining recognition from others. It focuses on the recognition and appreciation of individual achievements by peers or a wider audience [24]. Using artificially intelligent mobile fitness applications (MFAs), users can disseminate their data and achievements, thereby facilitating the expansion of their social networks. Disseminating fitness-related data among individuals with similar interests has the potential to strengthen social bonds and provide individuals with the motivation they need to achieve their health goals. Furthermore, these platforms facilitate competitive interactions among members, which in turn keeps users involved in their respective communities [32]. Sharing features allow users to share activity data, monitor personal social networking interactions and achievements, communicate, form chat groups, and organize events with friends [29]. Life coaches and physiotherapists at the Artificial Intelligence Multifunctional Forum can assist users in changing their exercise regimens, improving diets, avoiding injuries, and adopting healthy habits. Overall, Evidence of higher social engagement levels has been shown to positively affect users' commitment to exercise routines.

Physical goals. [18] noted that exercise has health benefits such as strength and endurance training, improved appearance, weight management, and overall management [18]. In fitness apps that focus on behavior change techniques, it is crucial to tailor goals to the user's specific situation, taking into account factors such as motivation and progress [26]. Goals need to be effective and precise. Make sure users can achieve their goals through some challenges that motivate learners by requiring effort without causing repeated failure. Goals must be precise, which means having a clear and targeted goal and setting it for the near term to ensure prompt commitment and action [8, 20]. In the absence of a professional coach, reinforcement learning algorithms allow a virtual coach to set and modify daily fitness goals based on the user's current and expected performance. This approach has been shown to increase motivation and self-efficacy, as well as the likelihood of achieving and exceeding goals [26]. Behr et al. [8] studied the impact of an AI mobile fitness app on weight management from an emotional dimension in a mobile weight loss program. The results showed that people who used more negative emotional language lost more weight, which suggests that it is important for AI-enabled mobile fitness apps to enhance tools to help users deal with negative emotions during goal setting and effort. This also shows that there is some interconnected relationship between different goals.

4.3 Outcomes

AI-powered mobile fitness apps offer users powerful tools to boost their health and well-being. Evidence suggests that such apps have a positive impact on many aspects of users' lives. These apps use AI to provide users with a personalised, adaptable, and engaging experience that delivers immediate and long-term benefits. Fitness software collects user information in the following ways: automatic collection by sensors and manual input by users. Automatic data collection is achieved by syncing and importing data from other applications or accessing built-in technologies or developer frameworks (such as HealthKit on iOS or Google Fit on Android). Then, machine learning algorithms or computing technologies are used to automatically analyze the data to provide valuable feedback and personalized services [2].

Goal attainment. Goal achievement can be supported through various settings, such as goal setting and technology backup. Goal setting (e.g., trailing feedback and goal-tracking features) is very important and has been installed on various mobile devices [41]. Aldossari [33] showed that system quality and information quality are two key determinants of MFA goal setting and goal tracking use, which have a significant impact on users' goal achievement and physical activity behavior changes. Goal setting and tracking are strongly associated with goal achievement and behavior change. In addition, positive reinforcement, such as praise, self-selected rewards, and immediate support, including providing tailored feedback [4]. Kuru demonstrated the effectiveness of sharing as emotional support, which can reduce negative emotional states and increase goal achievement. AI algorithms connect users with similar people, and the social aspect promotes an environment that increases motivation and accountability. In addition, studies conducted by Hietbrink et al. [16, 19], Kim, and Truelove et al. support these findings. The AI component appears to play a key role in setting realistic goals, tracking progress, and providing personalized recommendations that respond to users' changing needs and abilities.

Health outcome. AI-powered fitness apps offer personalized services to enhance users' health and well-being, using behavior change techniques (BCTs). BCTs are effective tools for sustainable behavior change that contributes to improved health outcomes. Kuru noted that goal setting, action planning, behavioral self-monitoring, and social support were among the top practices that significantly increased user engagement. Asbjornsen et al. concluded similar findings. [17] noted that prolonged use of fitness apps had long-term beneficial impacts on health and well-being beyond just weight loss, including reduced emotional exhaustion and enhanced overall quality of life satisfaction. Furthermore, making full use of fitness apps was found to significantly help users adhere to exercise regimes more successfully and experience increased overall satisfaction with overall quality of life satisfaction.

Habit formation. AI-powered fitness software has the power to facilitate habit formation among its users through guidance and reinforcement. Studies conducted by [16, 29, 32] highlight coaching elements where AI may play an essential part, providing personalized exercise plans, nutritional advice, lifestyle recommendations tailored specifically for individual user profiles, reinforcement features like adaptive reminders, reward systems or progressive challenges can encourage consistent behavior leading to healthy habits formation.

Continuance intention. Continued use intention is an indicator of app success [20, 32], and Kim (2021) have all conducted studies showing that users of AI-enabled fitness apps show high continued use intention. Lee and Lin [20] demonstrated that AI features such as intelligence and anthropomorphism can influence MFA users' continued use intention through goal difficulty and specificity. Goal specificity increases users' continued use intention, while goal difficulty decreases users' continued use intention. This may be attributed to the personalized experience, ongoing benefits, and how the app adjusts itself to the user's changing needs and goals. AI mobile fitness apps represent a significant advancement in mobile health technology. By breaking through the limitations of traditional fitness software and improving personalized and targeted fitness approaches, AI fitness apps create sustained engagement, promote health, and support people's pursuit of improved health and fitness in a digital form.

4.4 Challenges of MFA(AI)

The following three perspectives of the section discuss challenges associated with AI-enabled mobile fitness applications. The first issue is related to privacy issues and data security. The implementation of AI in mobile fitness apps presents some challenges, as supported by recent research. Privacy concerns and security issues related to self-tracking and monitoring emerge as primary obstacles, with studies by [2, 9, 11] highlighting the delicate balance between enhanced security and user privacy. The necessity for extensive data collection raises questions about the storage, access, and potential misuse of the data. The second issue focuses on Overload. the overload problem in the form of an overabundance of personalization options or reminding [11]. Personalization, while promising in theory, is limited by the specificity of feedback, specifically, some users find too many options, which can impair the system's ability to effectively adapt to individual user behaviors and preferences. This overload may cause user fatigue. The third challenge focuses on cost. Kuru investigated users' cost choices, with a minority of users believing the app's price was too high. Some users tried the cheaper option, which points to the need to more clearly inform users of the financial value of the app.

5 Limitations and Future Research

It is important to note that this systematic literature review of AI mobile fitness applications and goal attainment has some limitations. These limitations indicate potential directions for future research. Firstly, this review may be limited by language bias. Since this review is primarily based on studies published in English, it may have excluded relevant studies in other languages. It would be desirable for future research to be conducted without language limitations, thereby increasing the generalizability of the findings. Secondly, the relationship between AI features and goal attainment is further investigated. Thus, future research could investigate the impact of AI features on the user's goal achievement process, thereby advancing our understanding of goal achievement in AI-driven fitness software. Additionally, the accelerated development of AI and fitness technologies may render the existing literature rapidly obsolete and thus require constant updating to remain current. Finally, this review's reliance on existing literature may result in the omission of emerging trends or new applications that have not yet been extensively researched or published, thus limiting the scope of its findings. Examining the long-term impact of AI on goal attainment, habit formation, and intent to continue using, or how AI may shape the future role of coaches, could be considered an area for future research. Comparing the role of AI in various mobile fitness apps to other fields or industries to provide a broader perspective could also be considered scope for future research.

6 Conclusion

The role that AI mobile fitness apps have played in the success of the health and fitness sector cannot be underestimated. These applications are changing people's attitudes and approaches to health and fitness activities and goal achievement, providing personalized services through natural language processing and artificial intelligence data mining and predictive algorithms. The artificial intelligence algorithms analyse user data, such as physical activity, nutrition, and sleep indicators, to provide feedback and suggestions on how to best achieve goals and maintain progress while achieving desired goals.

AI-driven mobile fitness apps play an invaluable role in successfully achieving health and fitness goals. Their power lies in providing personalized assistance using technologies such as natural language processing and big data analytics, changing people's mindsets about maintaining health and fitness habits. For example, based on data such as a user's calorie consumption, sleep quality, and exercise activity, AI fitness apps can provide personalized recommendations to keep users healthy and achieve their fitness goals without risk.

The other point about AI apps is that they focus on the same social support. By sharing improvements and obtaining advice from the same people, people share a social group, so users can be motivated. Social elements such as a group's teamwork, competing as teams to finish a task or the use of leaderboards can help individuals to experience camaraderie. These social elements will help people to become emotionally connected, which is crucial to the realization of permanent fit lifestyles. As artificial intelligence is in the process of becoming more and more integrated into fitness technology, the relations between the developers and the users must be as balanced as possible.

Future researchers may try to address some of the shortcomings. Providing the same methods in other languages and pooling studies conducted in multiple languages can help to better understand the case. The literature should be updated because AI is a field that is constantly and rapidly changing and it is crucial to discuss its latest advances and any future trends. Longitudinal studies are encouraged because they provide long-term insights.

References

1. Abbass, H. (2021). What is artificial intelligence?. *IEEE Transactions on Artificial Intelligence*, *2*(2), 94–95.
2. Alwakeel, L., & Lano, K. (2022). Functional and technical aspects of self-management mHealth apps: systematic app search and literature review. *JMIR Human Factors*, *9*(2), e29767.
3. Alzahrani, A. S., Gay, V., Alturki, R., & AlGhamdi, M. J. (2021). [Retracted] Towards Understanding the Usability Attributes of AI-Enabled eHealth Mobile Applications. *Journal of Healthcare Engineering*, *2021*(1), 5313027.
4. Asbjørnsen, R. A., Wentzel, J., Smedsrød, M. L., Hjelmesæth, J., Clark, M. M., Solberg Nes, L., & Van Gemert-Pijnen, J. E. (2020). Identifying persuasive design principles and behavior change techniques supporting end user values and needs in eHealth interventions for long-term weight loss maintenance: qualitative study. *Journal of medical Internet research*, *22*(11), e22598.
5. Bagozzi, R. P., & Dholakia, U. (1999). Goal setting and goal striving in consumer behavior. *Journal of marketing*, *63*(4_suppl1), 19–32.
6. Bagozzi, R. P., & Warshaw, P. R. (1990). Trying to consume. *Journal of consumer research*, *17*(2), 127–140.
7. Beldad, A. D., & Hegner, S. M. (2018). Expanding the technology acceptance model with the inclusion of trust, social influence, and health valuation to determine the predictors of German users' willingness to continue using a fitness app: A structural equation modeling approach. *International Journal of Human–Computer Interaction*, *34*(9), 882–893.
8. Behr, H., Ho, A. S., Mitchell, E. S., Yang, Q., DeLuca, L., & Michealides, A. (2021). How do emotions during goal pursuit in weight change over time? Retrospective computational text analysis of goal setting and striving conversations with a coach during a mobile weight loss program. *International Journal of Environmental Research and Public Health*, *18*(12), 6600.
9. Bhargava, Y., & Nabi, J. (2020). The opportunities, challenges and obligations of Fitness Data Analytics. *Procedia Computer Science*, *167*, 1354–1362.
10. Cheng, C. H., & Chen, C. H. (2018). Developing a Mobile APP-Supported Learning System for Evaluating Health-Related Physical Fitness Achievements of Students. *Mobile Information Systems*, *2018*(1), 8960968.
11. Chew, H. S. J. (2022). The use of artificial intelligence–based conversational agents (Chatbots) for weight loss: scoping review and practical recommendations. *JMIR Medical Informatics*, *10*(4), e32578.
12. Corbin, C. B., Pangrazi, R. P., & Franks, B. D. (2000). Definitions: Health, fitness, and physical activity. *President's Council on Physical Fitness and Sports Research Digest*.
13. Farrokhi, A., Farahbakhsh, R., Rezazadeh, J., & Minerva, R. (2021). Application of Internet of Things and artificial intelligence for smart fitness: A survey. *Computer Networks*, *189*, 107859.
14. GSMA. (2023). The State of Mobile Internet Connectivity Report 2020 - Mobile for Development. Retrieved from Mobile for Development website: https://www.gsma.com/r/somic/
15. Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., ... & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. *Current reviews in musculoskeletal medicine*, *13*, 69–76.

16. Hietbrink, E. A., Middelweerd, A., van Empelen, P., Preuhs, K., Konijnendijk, A. A., Nijeweme-d'Hollosy, W. O., ... & Vollenbroek-Hutten, M. M. (2023). A digital lifestyle coach (E-Supporter 1.0) to support people with type 2 diabetes: participatory development study. *JMIR human factors*, *10*(1), e40017.
17. Hu, J., He, W., Zhang, J., & Song, J. (2023). Examining the impacts of fitness app features on user well-being. *Information & Management*, *60*(5), 103796.
18. James, T., Bélanger, F., & Lowry, P. B. (2021). The mediating role of fitness technology enablement of psychological need satisfaction and frustration on the relationship between goals for fitness technology use and use outcomes. *Journal of the Association for Information Systems (JAIS)*, *23*(4), 913–965.
19. Kramer, J. N., Künzler, F., Mishra, V., Presset, B., Kotz, D., Smith, S., ... & Kowatsch, T. (2019). Investigating intervention components and exploring states of receptivity for a smartphone app to promote physical activity: protocol of a microrandomized trial. *JMIR research protocols*, *8*(1), e11540.
20. Lee, J. C., & Lin, R. (2023). The continuous usage of artificial intelligence (AI)-powered mobile fitness applications: the goal-setting theory perspective. *Industrial Management & Data Systems*, *123*(6), 1840–1860.
21. Li, Y., Carter, M., & Crossler, R. (2023). Fitness Technology and Exercise Engagement: How Technology Affordances Facilitate Fitness Goal Attainment.
22. Liu, X., & Han, L. (2022). [Retracted] Artificial Intelligence Enterprise Management Using Deep Learning. *Computational Intelligence and Neuroscience*, *2022*(1), 2422434.
23. Liu, Y., & Avello, M. (2021). Status of the research in fitness apps: A bibliometric analysis. *Telematics and Informatics*, *57*, 101506.
24. Markland, D., & Ingledew, D. K. (1997). The measurement of exercise motives: Factorial validity and invariance across gender of a revised Exercise Motivations Inventory. *British Journal of Health Psychology*, *2*(4), 361–376.
25. Marr, B. (2018). What is artificial intelligence and how will it change our world.
26. Mohan, S. (2021). Exploring the role of common model of cognition in designing adaptive coaching interactions for health behavior change. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *11*(1), 1–30.
27. Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: the PRISMA Statement. *PLoS Medicine*, *6*(7).
28. Nouri, R., R Niakan Kalhori, S., Ghazisaeedi, M., Marchand, G., & Yasini, M. (2018). Criteria for assessing the quality of mHealth apps: a systematic review. *Journal of the American Medical Informatics Association*, *25*(8), 1089–1098
29. Oc, Y., & Plangger, K. (2022). GIST do it! How motivational mechanisms help wearable users develop healthy habits. *Computers in Human Behavior*, *128*, 107089.
30. Pires, I. M., Marques, G., Garcia, N. M., Flórez-Revuelta, F., Ponciano, V., & Oniani, S. (2020). A research on the classification and applicability of the mobile health applications. *Journal of personalized medicine*, *10*(1), 11.
31. Pressman, S. D., & Cohen, S. (2005). Does positive affect influence health?. *Psychological bulletin*, *131*(6), 925.
32. Puri, S., Pandey, S., & Chawla, D. (2022). Impact of technology, health and consumer-related factors on continued usage intention of wearable fitness tracking (WFT) devices. *Benchmarking: An International Journal*, (ahead-of-print).
33. Q. Aldossari, M., Nguyen, Q. N., Ta, A., & Schulz, S. A. (2023). System and Information Qualities in Mobile Fitness Apps and Their Effects on User Behavior and Performance. *Information Systems Management*, *40*(3), 270–289.
34. Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6), 420.
35. Sehar, U., & Naseem, M. L. (2022). How deep learning is empowering semantic segmentation: Traditional and deep learning techniques for semantic segmentation: A comparison. *Multimedia Tools and Applications*, *81*(21), 30519–30544.

36. Shabir, H., D'Costa, M., Mohiaddin, Z., Moti, Z., Rashid, H., Sadowska, D., ... & Cox, B. (2022). The barriers and facilitators to the use of lifestyle apps: a systematic review of qualitative studies. *European journal of investigation in health, psychology and education*, *12*(2), 144–165.
37. Shilts, M. K., Horowitz, M., & Townsend, M. S. (2004). Goal setting as a strategy for dietary and physical activity behavior change: a review of the literature. *American Journal of Health Promotion*, *19*(2), 81–93.
39. Truelove, S., Vanderloo, L. M., Tucker, P., Di Sebastiano, K. M., & Faulkner, G. (2020). The use of the behaviour change wheel in the development of ParticipACTION's physical activity app. *Preventive Medicine Reports*, *20*, 101224.
40. Yan, M., Filieri, R., Raguseo, E., & Gorton, M. (2021). Mobile apps for healthy living: Factors influencing continuance intention for health apps. *Technological Forecasting and Social Change*, *166*, 120644.
41. Zhou, M., Fukuoka, Y., Mintz, Y., Goldberg, K., Kaminsky, P., Flowers, E., & Aswani, A. (2018). Evaluating machine learning–based automated personalized daily step goals delivered through a mobile phone app: Randomized controlled trial. *JMIR mHealth and uHealth*, *6*(1), e9117.

Artificial Intelligence in Energy Informatics: An In-Depth Review

Safa Oleiwi, Saraswathy Shamini Gunasekaran, Moamin A. Mahmoud, Jaspaljeet Singh Dhillon, Nazirul Nazrin Shahrol Nidzam, and Salama Mostafa

Abstract AI is a new technology, and although it is seen by many in the energy business as a revolutionary one, there are some unknowns surrounding its influence. In the subsequent chapter, there will be a full discussion of the AI applications in the energy sector that are based on machine learning, deep learning, optimization techniques, and natural language processing. We investigate in depth the whole spectrum of the employment of those technologies, which span the domains of energy forecasting, demand side management, and ultimately anomaly detection, as well as additional fashionable phenomena. As far as each renewable energy field is concerned, provided are real-world examples to illustrate how AI may be utilized in the integration of renewable energy, smart grid management, energy efficiency, and energy market forecasts. In addition, we bring up challenges like data quality, inter-connectivity, and ethical dilemmas, as well as the future prospects of AI-based energy informatics trends. The objective of this study is to analyze AI applications in the energy sector presently and examine their influence on the environment to supply policymakers, researchers, and practitioners with knowledge in connection with artificial intelligence in a changing energy sector.

Keywords Artificial intelligence · Energy informatics · Machine learning · Deep learning · Optimization algorithms · Natural language processing · Renewable energy integration · Smart grid management · Energy efficiency · Energy market forecasting · Data quality · Interoperability · Ethical implications · Emerging trends

S. S. Gunasekaran · M. A. Mahmoud · J. S. Dhillon
College of Computing and Informatics, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia

Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia

S. Oleiwi (✉) · N. N. S. Nidzam
College of Graduate Studies, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia
e-mail: safahussain36@gmail.com

S. Mostafa
Faculty of Computing, Universiti Tun Hussein Onn, Parit Raja, Johor, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_11

1 Introduction

Information and energy have become two of the most vital components of human life. With technological advancement, informatics of energy has garnered interest in specific groups. This newly established industry makes use of the newest breakthroughs in the field of AI and IoT to bring forth not only the efficiency but also the dependability and sustainability of the energy infrastructure [1, 2]. The objective of this post is to study the possibilities of AI tools for recognizing and simplifying processes in the energy sector.

The spectrum of subjects, which include AI tools and energy information, is one of the priorities in the evaluation and covers machine learning, deep learning, optimization algorithms, and natural language processing [3, 4]. This talk will be focused on understanding the many approaches that are being utilized for tasks including energy forecasting, demand-side management, renewable energy integration, and smart grid optimization [5, 6].

The society brilliantly displays the relevance of this topic by displaying the requirement to transition towards sustainable energy systems when concerns about climate change and limited resources are emerging. Fast improvements in AI applications in recent times might be seen as a certain potential for the removal of environmental consequences through smart energy management employing AI driven solutions [7]. Nevertheless, the challenges are not going to be insignificant, including data quality problems, interoperability assumptions, and the ethical implications of implementing AI in the energy sector [8, 9]. By resolving these flaws and assigning AI to its position, we will attain the aim of making the future of energy renewal more dependable and eco-friendly [10]

We aim here at constructing a bridge among the current research, presenting the findings, pointing out the gaps in the literature, and also offering future avenues for applying AI in energy informatics. We are to emphasize and make public the technological tandem in the energy system so that we are able to contribute to the wider conversation of utilizing technology to find solutions to managing global energy concerns.

2 Background and Significance

2.1 Energy Informatics

Energy informatics is a cross-disciplinary discipline with contributions from energy systems, information technology, and data science. It ensures a well-functioning smart grid system [11]. At the core of Energy Informatics lies the application of digital technology to boost the efficiency of operations, including production, delivery, and usage of energy. The energy informatics section of the study will put forward the essential components, including its aims, problems, and practical applications, backed by academic resources [12].

Goals: The fundamental objective of energy informatics is to create the conditions for a seamless change of the existing energy systems to be more sustainable and efficient. Using a range of computational methodologies and data analysis tools, energy informatics strives to increase energy efficiency [13], minimize waste, and boost the integration of renewable energy resources (RESs) into grid operations [14, 15]. Furthermore, the goal of energy informatics is to deliver power to customers through continuous energy data and insights for wiser and more informed decision-making [16].

Obstacles: Energy informatics may hold a lot of potential benefits, but it also comes with certain obstacles that might limit its application on a large scale. For instance, data interoperability challenges and cybersecurity concerns are part of these obstacles, combined with the complexity of the integration of different energy systems and technologies [17, 18]. In addition, the hastened rate of technological innovation is producing issues that are largely connected to a gap in skills and worker preparation [19].

Applications: Energy data science employs varied devices in smart grid control, machine learning, building energy management, and electric automobiles [15, 20]. The purpose of the figure is to analyze the different energy informatics' uses, including the success of their use in energy usage optimization and in boosting power system efficiency.

What we see is that informatics in the energy sector provides a vital aspect of integrating the current energy infrastructure with renewable energy sources, for instance, wind and solar [21, 22]. In Fig. 2, we study the problems of solar integrations in isolated industrial microgrids, where energy visualization- tools such as energy informatics are vital to minimizing dependability concerns and boosting overall system performance [23].

To finish, energy information can be considered one of the most beneficial ways to increase energy sustainability and effectiveness. However, energy informatics has a tremendous opportunity for advancement, as its objective is to handle the difficulties while leveraging the latest technology. This presents an opportunity for energy informatics to act as a driver of improvements and to be incorporated in worldwide efforts to reach a more sustainable future.

2.2 Role of AI

It cannot be denied that AI is a significant aspect of enhancing efficiency and striving to solve the challenges that come up during data collection in energy informatics. Through the leveraging of AI technology, energy systems are able to increase energy system efficiency, dependability, and sustainability in numerous domains. Figure 3 A schematic representations illustrating the role of AI in optimizing energy processes

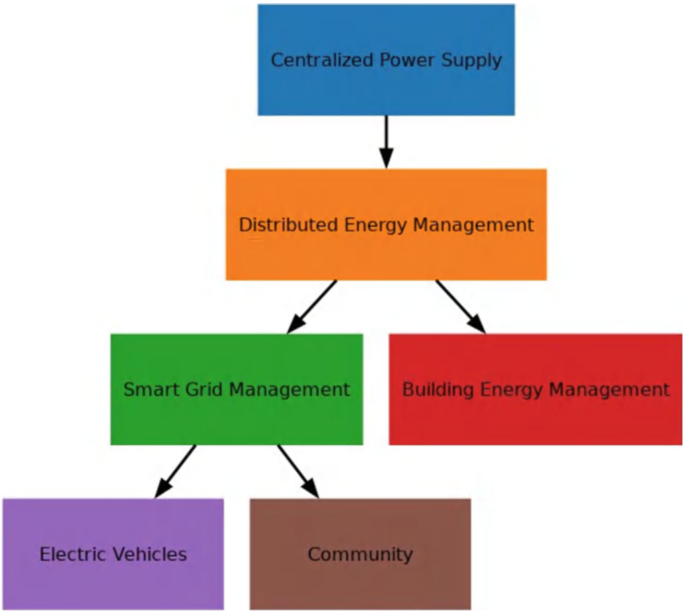


Fig. 1 A schematic representation of the key application areas of energy informatics, including smart grid management, industrial automation, building energy management, and transportation electrification

within energy informatics, highlighting the integration of machine learning, IoT, and graph signal processing techniques.

Machine Learning: Machine learning algorithms are a property of the energy systems that enable them to do jobs that will analyze vast quantities of data and extract insights that will help optimize operations [22, 24]. Machine learning algorithms may be utilized to estimate the downtime and maintenance costs of equipment using maintenance models with predictive algorithms that predict equipment faults [25, 26].

The Internet of Things (IoT): The aggregation of intelligent IoT devices with AI provides real-time monitoring as well as management of energy systems, with the beneficial effect of greater energy efficiency and resource utilization [27, 28]. IoT-based smart grids integrate AI algorithms, which allow regulating the energy distribution, balancing the load, and integrating renewable sources' [29, 30].

Graph Signal Processing: Through processing graph signal approaches, one may examine a complicated energy distributed system to achieve optimal flow and distribution [31, 32]. An apparent conclusion is that when energy systems are seen as graphs, the AI algorithms may discover the most effective paths to reduce energy losses [33, 34].

Obstacles and Opportunities: AI in energy informatics is building up expectations, yet the obstacles to be mastered to utilize it to its maximum potential may emerge as a barrier [35]. Interconnection challenges involve data incompatibility, cybersecurity

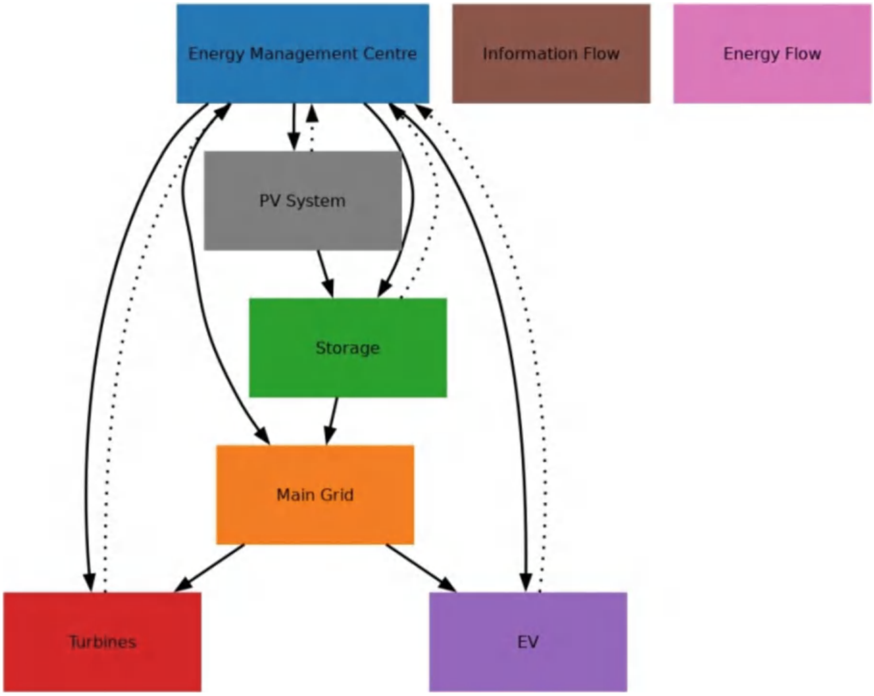


Fig. 2 An illustration depicting the challenges of solar power integration in isolated industrial microgrids, highlighting the role of energy informatics in overcoming reliability constraints and optimizing system performance

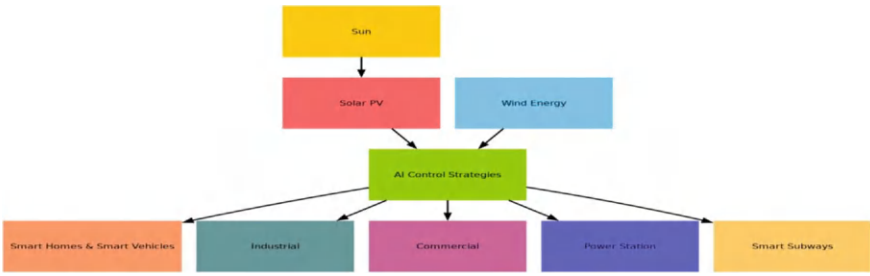


Fig. 3 A schematic representation illustrating the role of AI in optimizing energy processes within energy informatics, highlighting the integration of machine learning, IoT, and graph signal processing techniques

problems, and the complexity of merging multimodal energy systems [36]. Another component is two-spiritedness [29], which indicates that the individual is genuinely two-spirit when it does not simply point to the gender or the orientation of the person but also to his soul and heart. Table 1 represent the key AI techniques and

Table 1 Key AI techniques and their applications in energy informatics

AI technique	Applications in energy informatics
Machine learning	Predictive maintenance, energy forecasting, anomaly detection
Internet of Things (IoT)	Smart grid management, real-time monitoring, energy consumption optimization
Graph signal processing	Energy flow optimization, network analysis, fault detection

their applications in energy informatics, including machine learning, IoT, and graph signal processing.

2.3 Previous Research

The integration of artificial intelligence (AI) applications in energy informatics has been one of the primary issues of interest in the academic sector, and the researchers have been introducing the technical potentials, barriers, and expansions here. This sub-section outlines the already recognized concerns via extensive evaluation of current works and illustrates which experimental design is needed for studies that are not yet extant.

Watson et al. (2010) [6], in their study, mainly underlined that "energy informatics" has been designed to bridge the gap in the preservation of the environment utilizing data systems [37]. This famous paper emphasized the usage of information technology to cope with energy-related challenges and gave fresh perspectives for the information systems discipline to contribute to the goals of a cleaner and ecologically sustainable environment. See Fig. 4 for an illustration of smart energy system modeling.

Andreopoulou (2012) broadened the notion of "green informatics" in the context of applying ICT (Information and Communication Technology) to advancing ESD (Environmental Sustainability Development [37]). When this research was done, it was recognized that ICT systems and tools might be effective in enhancing efforts that are green in a diversity of domains, such as energy management and conservation.

Zhou and Yang (2016) illustrated how big data technologies might be utilized to study home energy usage behavior [38]. The study has been feasible through a process of evaluating the enormous amounts of data that have reflected patterns and elements that contribute to how energy is used, therefore revealing areas for energy consumption enhancement and promotion of energy efficiency at the home level [39].

Palensky and Dietrich (2011) investigated the demand side management area of interest within their research, where they extensively evaluated demand response, intelligent energy systems, and smart loads [40]. Such a study has demonstrated the function of demand-side management in optimizing energy consumption patterns,

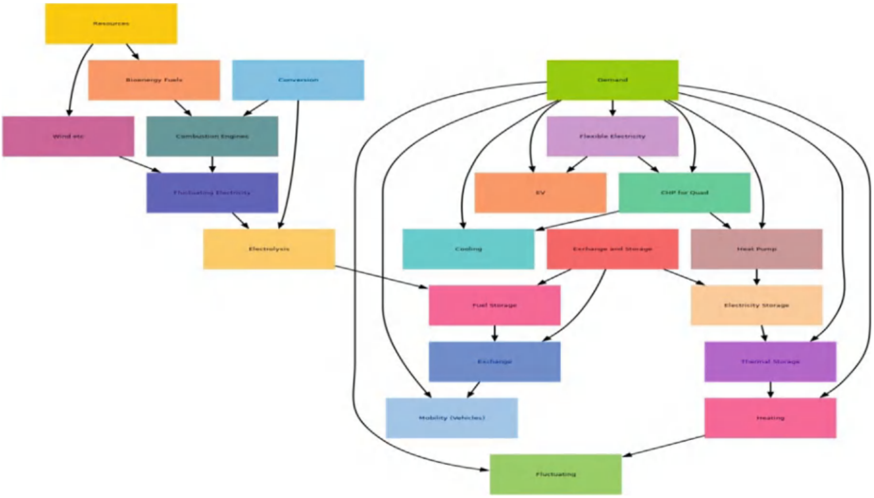


Fig. 4 Smart energy system modeling

producing more stable grids, and bringing about efficiency in resource allocation [27].

Rehmani et al. (2018) analyzed current achievements with respect to the integration of renewable energy sources into a smart grid, with an emphasis on the role of ICT [30]. This analysis stressed ICT infrastructure involvement in the effective winding up of the integration of renewable power sources, grid monitoring, and load adjustments [15].

Jiang et al. (2016) based their study on an analysis of big data in energy and pointed out the primary obstacles and possibilities in the existing quantity of data context [41]. Here, the primary data sources and analytic tools for energy big data, as well as trending management challenges, were discovered, which should pave the way for future research in this field [30, 42].

Whereas Usman and Shami (2013) discussed the growth of digital technologies for smart grid applications and assessed the improvements in protocols and networking technologies as well as their usefulness for smart grid implementation (Stricha, Zaccariotto, and Ponti, 2018 [43]). The study brought attention to the power of monitoring energy sources by employing communication infrastructure and their real-time coordination, control, and optimization.

As one of the most important contributions, Bordin et al. (2020) sought to explain smart energy and power systems modeling by exploring the roles of IoT and cyber-physical systems implementation at the energy informatics level [25]. This important work shed light on the intricacies of an energy system model, highlighting the incorporation of IOT technologies and cyber-physical frameworks for surveillance, control, and system optimization through increased monitoring. The study work of Bordin et al. will serve as the basis for comprehension of the current energy domain, including the potential of AI to contribute to energy solutions to conquer the issues.

In the current study, Saheb et al. (2022) applied the notion of contextual topic modeling and content analysis to evaluate the extent of AI engagement in sustainable energy driven initiatives [27]. The research has been beneficial in uncovering the celebratory trends, problems, and opportunities of utilizing AI as a tool in the campaign to support sustainability in the energy sector. Through the correct examination of varied sorts of literature, Saheb et al. exposed us to multiple examples of AI use in power management, where it is being utilized for optimal energy resource allocation or energy efficiency, finding the path to renewable energy sources.

The contribution of Bordin et al. (2021) to the topic regarding teaching future energy informatics professionals is the focus of the study, which provides their perspectives on the prospective problems and prospects of this emerging sector [3]. This work has stressed that multidisciplinary education and training programs are vital in the preparation of professionals to handle complex energy informatics concerns. It also emphasized the necessity to blend theory with practical application in the work that professionals undertake in the laboratory and the institution.

The article by Pant et al. in 2023 has carried the issue further by describing the probable linkages between AI and IoT, which are both extremely revolutionary technologies [4]. This inquiry sheds light on the prospect of AI integrated IoT solutions to the problem of lowering energy consumption, increasing the resilience of the power grid, and ways to become part of sustainable growth. Through the huge use of data analytics and real time monitoring, the artificial intelligence-driven IoT systems deliver enhanced answers to the issues of energy management in the green project.

Tomazzoli et al. (2023) have created an IoT and AI framework integrating energy effectiveness in smart settings, which allows smart management of energy, as stressed by [5]. It was brought to light throughout the research that intelligent sensing, data analysis, and adaptive control mechanisms stand as the major methods utilized for enhancing energy efficiency and minimizing waste when applied to limited domains such as smart buildings or industries (see Fig. 5 for an overview of IoT and AI integration in energy efficiency). Furthermore, Table 2 presents a summary of the key findings and contributions of the reviewed literature, providing a structured overview of the diverse research landscape in AI applications for energy informatics.

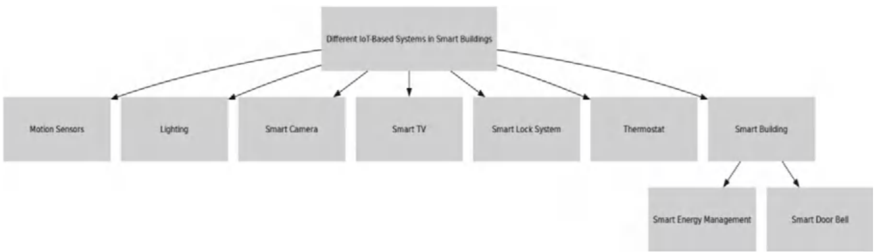


Fig. 5 Overview of IoT and AI integration in energy efficiency

Table 2 Summary of methodologies, techniques, findings, and limitations in AI applications for energy informatics

Author et al. (Reference number)	Methodology	Technique	Dataset name	Finding	Limitation
Bordin et al. [1]	Modeling	IoT and cyber-physical systems	Smart energy dataset	Enhanced system monitoring and control	Limited scalability of IoT infrastructure
Saheb et al. [2]	Topic modeling and content analysis	Artificial intelligence	Sustainable energy dataset	Identification of emerging trends and challenges	Dependency on the quality and diversity of input data
Bordin et al. [3]	Literature review	Educational strategies	Energy informatics curriculum dataset	Opportunities and challenges in educating specialists	Generalizability of findings may vary across educational contexts
Pant et al. [4]	Integration	AI and IoT	Green energy dataset	Optimization of energy consumption	Integration challenges in heterogeneous IoT environments
Tomazzoli et al. [5]	Analysis	IoT and AI	Energy efficiency dataset	Optimization of energy usage and waste reduction	Dependency on accurate sensor data and real-time analytics capabilities
Bakare et al. [11]	Overview and future directions	Demand side management	Smart grid dataset	Identification of challenges and solutions	Generalizability of findings may be limited by specific grid configurations
Das and Chandra [10]	Review	Fog computing	Fog computing dataset	Identification of challenges and potential applications	Variability in fog computing implementations and architectures
De Simone et al. [13]	Review	AI/ML in manufacturing	Manufacturing dataset	Identification of solved issues, limits, and challenges	Dependency on the availability and quality of manufacturing data

(continued)

Table 2 (continued)

Author et al. (Reference number)	Methodology	Technique	Dataset name	Finding	Limitation
Abir et al. [14]	Overview	IoT-enabled smart grid	Smart grid dataset	Applications and challenges in smart grid development	Scalability challenges in large-scale smart grid deployments
Mao et al. [16]	Overview	Industrial IoT	IIoT dataset	Overview of energy-efficient IIoT systems	Challenges in interoperability and standardization of IIoT protocols
Ma et al. [17]	Overview	Digitalization	Building-to-grid dataset	Integration of digitalization in building energy systems	Adaptability challenges in legacy building infrastructure
Zhang et al. [18]	Design and analysis	Informatics-aided design	Li metal anode dataset	Informatics-aided design of Li metal anode	Dependency on accurate material properties and computational models

While previous research has offered useful insights into the application of AI in energy informatics, various gaps and topics for additional inquiry remain. These include:

- Interdisciplinary Collaboration: Encouraging better collaboration across information systems, energy engineering, and environmental science fields to handle complex energy concerns.
- Ethical and Societal Implications: Investigating the ethical implications of AI-driven energy systems, including concerns of privacy, equality, and societal effect.
- Scalability and Interoperability: Exploring scalable AI solutions that can adapt to varied energy systems and ensuring interoperability across different AI technologies and energy infrastructure.
- Resilience and Security: Enhancing the resilience and security of AI-enabled energy systems against cyber attacks, physical weaknesses, and disturbances.
- By solving these gaps and obstacles, future research in AI applications in energy informatics might lead to more sustainable, efficient, and resilient energy systems (Table 2).

3 Bibliometric Analysis

Bibliometric analysis gives a thorough perspective of the research area by assessing publishing patterns, prominent journals, top authors, cooperation networks, keyword trends, and citation analysis. In this part, we report the findings of a bibliometric study done on the literature relevant to artificial intelligence (AI) in energy informatics. This study presents a macro picture of the research topic and emphasizes noteworthy trends and discoveries.

3.1 Methodology

Data for this investigation was acquired from reliable sources such as Web of Science, Scopus, and Google Scholar. We utilized search phrases like "artificial intelligence in energy informatics," "AI in sustainable energy," and "machine learning in energy systems" to obtain relevant literature. The acquired data was evaluated using bibliometric tools—VOSviewer and Bibliometrix (an R program).

3.2 Publication Trends

The year-wise distribution of publications linked to AI in Energy Informatics is illustrated in Figure. The data demonstrates a large rise in research papers over

Table 3 Top journals in AI and energy informatics

Rank	Journal name	Number of publications
1	Energy Informatics	25
2	IEEE Transactions on Industrial Informatics	18
3	Renewable and Sustainable Energy Reviews	15
4	Sustainable Computing: Informatics and Systems	12
5	Journal of Ambient Intelligence and Humanized Computing	10

Table 4 Key authors in AI and energy informatics

Rank	Author name	Number of publications	Total citations
1	C. Bordin	10	150
2	T. Saheb	8	120
3	S. Mishra	7	100
4	A. Håkansson	6	90
5	J. Singh	5	80

the previous decade, demonstrating increased interest and accomplishments in this sector.

3.3 Top Journals

Table lists the top journals publishing research in AI and Energy Informatics. These journals are pivotal in disseminating knowledge and advancements in the field (Table 3).

3.4 Key Authors

Table presents the leading authors in the field of AI and Energy Informatics, along with their contributions. These authors have significantly influenced research directions and advancements (Table 4).

3.5 Co-authorship Networks

Figure visualizes the co-authorship networks among researchers in the field. The network diagram highlights collaborative relationships and key contributors.

4 Keyword Analysis

Table lists the most common keywords in the literature, indicating the major themes and research focus areas (Table 5).

4.1 *Citation Analysis*

Table highlights the most cited papers and authors, showcasing influential works in the research domain (Table 6).

4.2 *Discussion*

The bibliometric study gives useful insights into the research area of AI in energy informatics. The rising publishing patterns show a growing interest and considerable progress in integrating AI with energy systems. Top publications such as Energy Informatics and IEEE Transactions on Industrial Informatics play a key role in spreading research results. Leading writers like C. Bordin and T. Saheb have considerably contributed to the area, and the co-authorship networks emphasize crucial collaboration links.

Table 5 Common keywords in AI and energy informatics

Rank	Keyword	Frequency
1	Artificial Intelligence	45
2	Energy Informatics	40
3	Sustainable Energy	35
4	Machine Learning	30
5	Internet of Things (IoT)	25

Table 6 Most cited papers in AI and energy informatics

Rank	Paper (Author, Year)	Citations
1	Watson et al. (2010), MIS Quarterly	300
2	Sisinni et al. (2018), IEEE Transactions on Industrial Informatics	250
3	Jiang et al. (2016), IEEE Access	200
4	Rehmani et al. (2018), IEEE Transactions on Industrial Informatics	180
5	Friedewald and Raabe (2011), Telematics and Informatics	150

The keyword analysis suggests that artificial intelligence, energy informatics, and sustainable energy are important topics, showing the emphasis on creating intelligent and sustainable energy solutions. The citation analysis reveals key articles and authors, giving a basis for future research initiatives.

This bibliometric study presents a macro picture of the research topic, showing major trends, prominent contributions, and cooperation patterns. These findings may drive future research and support greater improvements in AI and energy informatics.

5 Methodology

The given methodology approach for the literature analysis of the artificial intelligence (AI) function in energy informatics is outlined in this part. Our process comprises three critical components: a complete examination of papers, data extraction, and quality assessment. We have covered the period of a year, starting with the early search of the articles using recognized databases and relevant keywords, which have led us to the paper we uncover. Subsequently, data gathering was concentrated on synthesizing the fundamental information that was taken from the selected research, such as AI methodologies, application areas, datasets employed, important results, problems, and limits. The rigorous assessment of studies, including just the most reliable material and rejecting those that can be biased, was undertaken to assure a repeat of high-quality research. Through such a methodological framework, the review process is guided in relation to the definition of methodological rigor, comprehensiveness, and competency in examining state-of-the-art advances in AI applications in the energy sector. Figure 6 is an all-inclusive flowchart that the authors have prepared in order to offer an overview of the reviewed article. It is the flowchart that practically displays the order of events happening in completing the study of the literature search strategy, extracting information, making observations, and quality authentication. Specific actions of each phase of the review process are illustrated sequentially through the graph, exhibiting a systematic technique in analyzing the literature on AI involvement in smart energy systems. This chart, however, is merely a broad reference, meaning it does not provide data on the intricacies of the methodological framework employed in our study (Figs. 7 and 8).

6 Applications and Case Studies

6.1 Renewable Energy Integration

The intermittent renewable energy sources provide specific issues for the power system operators that execute the integration of these sources into the grid. Although AI has been applied as a solution to alleviate many of the expected issues and optimize

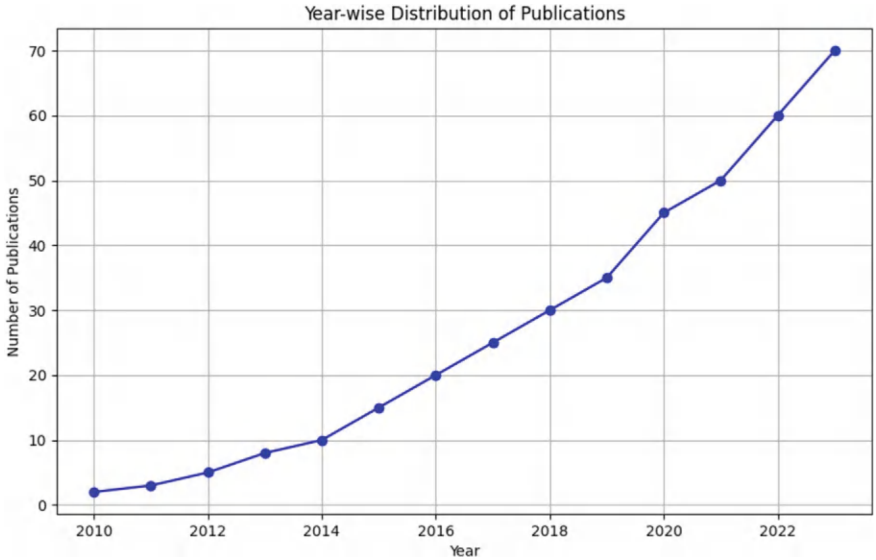


Fig. 6 Year-wise distribution of publications in AI and Energy Informatics

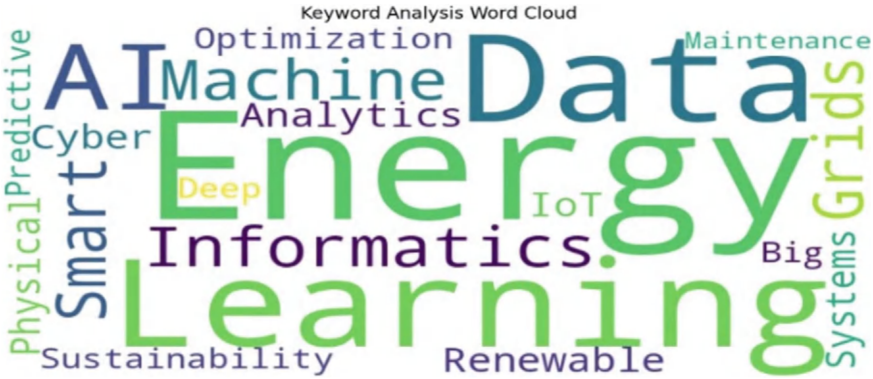


Fig. 7 Co-authorship networks in AI and energy informatics

Fig. 8 Overview of the reviewed article

integration, this technology may pose other concerns. Practical examples and case studies illustrate how AI is truly able to cope with the challenge of fulfilling the demands of renewable energy resources, which is to boost efficiency. An example illustrate in Fig. 9.

Case Study 1: A joint investigation between Energy Solutions and AI modules anticipated solar power generation utilizing historical weather information, data trends, and effective solar panel operating schemes. Through accurate forecasting of the solar output, operators of the grid are able to optimize the resources sufficiently, and thus, they are better equipped to allocate supply and demand more efficiently, which in turn boosts their usage of renewable energy sources.

Case Study 2: Wind Farm Optimization AI-based predictive maintenance has been deployed at ABC Renewables‘ wind farms to ensure the smooth functioning of their turbines and replace damaged components on a timely basis. AI processing may compare current sensor data with projected data alerts to identify an instance of equipment breakdown and conduct timely repair proactively, thereby lowering downtime and enhancing dependability for power generation.

6.2 Smart Grid Management

AI does not only aid but also plays a very crucial part in the optimization of smart grid activities such as efficient energy distribution, fault detection, and prediction of when an asset requires maintenance. The smart grids, with the assistance of modern analytics and machine learning algorithms, can act dynamically and, in this way, handle fluctuating demands and grid circumstances, thus ensuring trustworthy and stable functioning. Figure 10 show the example of Smart Grid Management.

Case Study 1: Dispatching and Demand Responses The AI-based models, coupled with the DEF Utility Company, were implemented in tandem, and they could properly estimate the electrical consumption trends. By employing prior data and consumption estimates, along with demographic information, smart grids are able to produce and redistribute electricity in real-time. This, in turn, optimizes load balancing and decreases peak demand.

Case Study 2: AI driven fault detection techniques applied by GHI Power Grids succeeded in isolating equipment failures and odd events in grid infrastructure. By applying analytics to the outputs from sensors, AI can identify early symptoms of equipment deterioration and schedule maintenance before any defect happens, which would reduce grid operation failures and assure grid dependability.

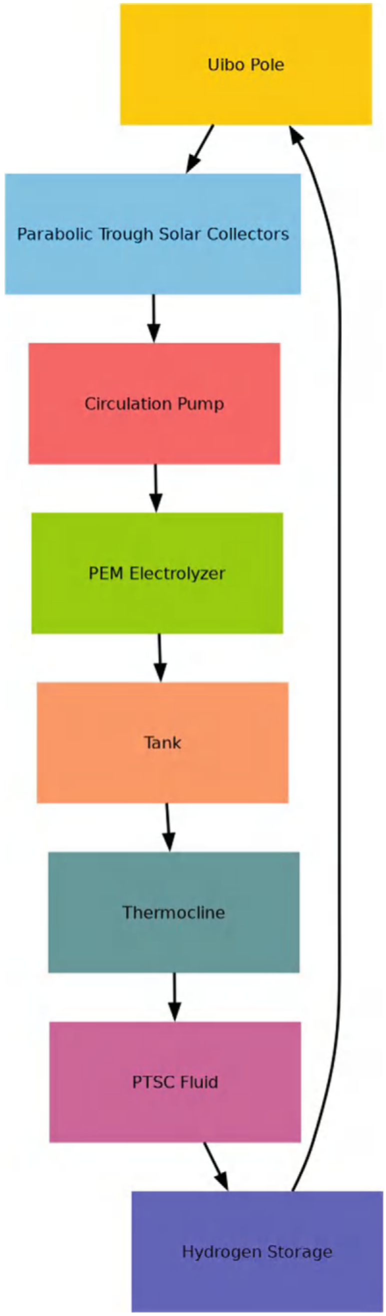


Fig. 9 Renewable energy integration

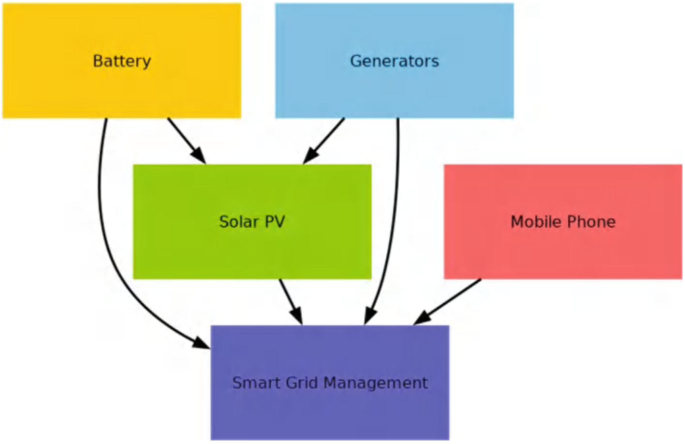


Fig. 10 Smart grid management

6.3 Energy Efficiency

The AI solution harnessed presents a huge promise of enhancing energy efficiency in many sectors, like buildings, industrial operations, and transportation, considerably better. Applying data analytics and optimization methodologies, AI specialize in integrating resources more efficiently and cutting down on energy usage and environmental effects.

Case Study 1: BEMS (Building Energy Management Systems) that are AI-based in partnership with JKL Corporation have been rolled out and are utilized to check into the energy use in commercial buildings. HVAC systems, lighting, and other energy-saving devices may be adjusted with the aid of AI that forecasts occupancy patterns, weather conditions, and general building dynamics. It results in energy, which is a large amount on the bills, and a reduction in expenditures.

Case Study 2: Industrial Process Tweaking MNO Manufacturing created an AI-based optimizer to boost invoicing efficacy in their production units. AI can achieve this by detecting process data and diagnosing inefficiencies, as well as improving machine operations. The consequence is the capacity to decrease energy use, restrict waste output, and boost industrial productivity, therefore taking the initiative for a sustainable manufacturing system.

6.4 Energy Market Forecasting

The employment of AI approaches is a crucial component in energy price forecasting, analyzing the patterns of demand in connection to the market as well as anticipating market volatility. This allows the stakeholders to make decisions with a clear head and utilize resources to the utmost ideal. AI systems can evaluate a large quantity of data on market trends and provide projections that can limit the risks of losing money and help them benefit from higher returns.

Case Study 1: AI-Driven Pricing Prediction in the Wholesale Electricity Market Jointly with PQR Power Traders, neural network-based models were built to allow data to "learn" the patterns of wholesale market price evolution and to be utilized subsequently in the automated power trading system. Based on historical market data, clouded skies predictions, and supply–demand dynamics, AI algorithms are able to predict price swings with high precision, which subsequently helps traders to accurately attune trading strategies and optimize earnings.

Case Study 2: Demand Forecasting for Energy Retailers STU has created AI-powered demand forecasting algorithms that are able to anticipate what and how much end-users will be consuming. The ability of AI, which is the appreciation of data from prior use, demographic aspects, and seasonal changes, provides exact projections of demand, and from this advantage, merchants can optimize pricing, better manage their inventory, and enhance customer pleasure.

As a consequence, the employment of AI systems in renewable energy integration, smart grid control, optimization of energy efficiency, and market automation will deliver an awesome future by helping the power industry to go up a notch in innovation, sustainability, and energy resource efficiency. The depiction of real-life case studies and examples stresses the revolutionary participation of AI in placing the energy sector into the new chapter of cleaner and more sustainable growth.

7 Challenges and Future Directions

Both the merging of AI and energy informatics could bring challenges as well. Discovering the answers to these issues as well as the areas of future study are vital conditions for the advancement of the technology and its rise in effectiveness in energy operations.

7.1 Data Quality and Availability

One of the primary stumbling points in the field of energy informatics is infinite data and its quality. Data sources in the energy market may come in numerous kinds and formats, as well as being either diversified or fractured. Nonetheless, such uncertainty still prevails, which is really a challenging test for data integration and data analysis. As the data becomes one of the most essential AI driving variables, the challenges linked to the quality of the data (including incompleteness, inconsistency, and errors) might reduce the efficacy of AI-deep driven applications. As with this solution, sound data governance structures, data quality concerns, and capabilities for gathering and integrating data will be needed. Besides, maintaining data safety and security top-notch is practically a precondition for any trust and faith in autonomously controlled energy systems to be implanted.

7.2 Interoperability

Interoperability continues to be a key hurdle to the widespread deployment of AI in the energy field. The absence of an agreed-upon data format, protocol, and interface may be the biggest hurdle to the seamless collaboration and interoperability of various energy systems and gadgets. In light of this dilemma, attempts to build industry-wide standards as well as interoperable systems where data exchange, intercommunication, and cooperation are feasible across various types of energy infrastructures that are interconnected. The standardization initiatives of the wider AI community are prone to rising interoperability, compatibility, and scalability levels, hence facilitating the integration of AI-based solutions in the progressive energy sector and gaining new benefits in the form of optimizations, developments, and innovations.

8 Conclusion

In essence, the introduction of artificial intelligence (AI) to energy informatics is akin to a new paradigm that bears huge ramifications. While AI is performed with the use of machine learning, deep learning, optimization algorithms, and natural language processing to assist in the in the search for novel solutions for the vast spectrum of energy-related difficulties, The practical examples of AI use in the fields of integrating renewable energy sources into grids, operating the smart grid, enhancing energy efficiency, and projecting market trends give proof of the effectiveness of AI technologies. The benefits include that AI can generate dependable and sophisticated models for anticipating energy consumption, but factors like data quality, interoperability, and ethical concerns need to be taken care of to attain the full potential of AI in business. However, another set of challenges arises in the context of the rising trends, like the interplay of AI with other new technologies. These new sectors provide prospects for future studies and technological improvements. Adopting and conquering these obstacles and welcoming transformative changes, AI-powered energy informatics may be the cutting-edge technology allowing viable, developed, and responsive energy.

References

1. Bordin, C., Håkansson, A., & Mishra, S. (2020). Smart energy and power systems modelling: an iot and cyber-physical systems perspective, in the context of energy informatics. Procedia Computer Science, 176, 2254–2263.
2. Saheb, T., Dehghani, M., & Saheb, T. (2022). Artificial intelligence for sustainable energy: A contextual topic modeling and content analysis. Sustainable Computing: Informatics and Systems, 35, 100699.
3. Bordin, C., Mishra, S., Safari, A., & Eliassen, F. (2021). Educating the energy informatics specialist: opportunities and challenges in light of research and industrial trends. SN Applied Sciences, 3, 1–17.
4. Pant, S., Rawat, P., Kathuria, S., Singh, R., Chanti, Y., & Pachouri, V. (2023, September). Artificial Intelligence and Internet of Things Intersection in Green Energy. In 2023 3rd International Conference on Innovative Sustainable Computational Technologies (CISCT) (pp. 1–5). IEEE.
5. Tomazzoli, C., Scannapieco, S., & Cristani, M. (2023). Internet of things and artificial intelligence enable energy efficiency. Journal of Ambient Intelligence and Humanized Computing, 14(5), 4933–4954.
6. Jarrahi, M. H., Lutz, C., Boyd, K., Oesterlund, C., & Willis, M. (2023). Artificial intelligence in the work context. Journal of the Association for Information Science and Technology, 74(3), 303–310.
7. Bordegoni, M., & Ferrise, F. (2023). Exploring the intersection of metaverse, digital twins, and artificial intelligence in training and maintenance. Journal of Computing and Information Science in Engineering, 23(6), 060806.
8. Tavares, D., Lopes, A. I., Castro, C., Maia, G., Leite, L., & Quintas, M. (2023). The intersection of artificial intelligence, telemedicine, and neurophysiology: opportunities and challenges. Handbook of Research on Instructional Technologies in Health Education and Allied Disciplines, 130–152.
9. Arun, A., Haque, M. M., Washington, S., & Mannering, F. (2023). A physics-informed road user safety field theory for traffic safety assessments applying artificial intelligence-based video analytics. Analytic methods in accident research, 37, 100252.
10. Das, K. P., & Chandra, J. (2023). A survey on artificial intelligence for reducing the climate footprint in healthcare. Energy Nexus, 9, 100167.
11. Bakare, M. S., Abdulkarim, A., Zeeshan, M., & Shuaibu, A. N. (2023). A comprehensive overview on demand side energy management towards smart grids: challenges, solutions, and future direction. Energy Informatics, 6(1), 4.
12. Das, R., & Inuwa, M. M. (2023). A review on fog computing: Issues, characteristics, challenges, and potential applications. Telematics and Informatics Reports, 10, 100049.
13. De Simone, V., Di Pasquale, V., & Miranda, S. (2023). An overview on the use of AI/ML in Manufacturing MSMEs: solved issues, limits, and challenges. Procedia Computer Science, 217, 1820–1829.
14. Abir, S. A. A., Anwar, A., Choi, J., & Kayes, A. S. M. (2021). Iot-enabled smart energy grid: Applications and challenges. IEEE access, 9, 50961–50981.
15. Wang, D., Zhong, D., & Souri, A. (2021). Energy management solutions in the Internet of Things applications: Technical analysis and new research directions. Cognitive Systems Research, 67, 33–49.
16. Mao, W., Zhao, Z., Chang, Z., Min, G., & Gao, W. (2021). Energy-efficient industrial internet of things: Overview and open issues. IEEE transactions on industrial informatics, 17(11), 7225–7237.
17. Ma, Z., Clausen, A., Lin, Y., & Jørgensen, B. N. (2021). An overview of digitalization for the building-to-grid ecosystem. Energy Informatics, 4(Suppl 2), 36.
18. Zhang, L. S., Gao, X. L., Liu, X. H., Zhang, Z. J., Cao, R., Cheng, H. C., ... & Yang, S. C. (2022). CHAIN: unlocking informatics-aided design of Li metal anode from materials to applications. Rare Metals, 41(5), 1477–1489.

19. Polleux, L., Guerassimoff, G., Marmorat, J. P., Sandoval-Moreno, J., & Schuhler, T. (2022). An overview of the challenges of solar power integration in isolated industrial microgrids with reliability constraints. Renewable and Sustainable Energy Reviews, 155, 111955.
20. Sherazi, H. H. R., Zorbas, D., & O'Flynn, B. (2022). A comprehensive survey on RF energy harvesting: Applications and performance determinants. Sensors, 22(8), 2990.
21. He, P., Almasifar, N., Mehbodniya, A., Javaheri, D., & Webber, J. L. (2022). Towards green smart cities using Internet of Things and optimization algorithms: A systematic and bibliometric review. Sustainable Computing: Informatics and Systems, 36, 100822.
22. Wu, Y., Wu, Y., Guerrero, J. M., & Vasquez, J. C. (2021). A comprehensive overview of framework for developing sustainable energy internet: From things-based energy network to services-based management system. Renewable and Sustainable Energy Reviews, 150, 111409.
23. Hu, Y., & Man, Y. (2023). Energy consumption and carbon emissions forecasting for industrial processes: Status, challenges and perspectives. Renewable and Sustainable Energy Reviews, 182, 113405.
24. Yu, Z. J., Haghighat, F., & Fung, B. C. (2016). Advances and challenges in building engineering and data mining applications for energy-efficient communities. Sustainable Cities and Society, 25, 33–38.
25. Khatua, P. K., Ramachandaramurthy, V. K., Kasinathan, P., Yong, J. Y., Pasupuleti, J., & Rajagopalan, A. (2020). Application and assessment of internet of things toward the sustainability of energy systems: Challenges and issues. Sustainable Cities and Society, 53, 101957.
26. Prieto, A., Prieto, B., Ortigosa, E. M., Ros, E., Pelayo, F., Ortega, J., & Rojas, I. (2016). Neural networks: An overview of early research, current frameworks and new challenges. Neurocomputing, 214, 242–268.
27. Sisinni, E., Saifullah, A., Han, S., Jennehag, U., & Gidlund, M. (2018). Industrial internet of things: Challenges, opportunities, and directions. IEEE transactions on industrial informatics, 14(11), 4724–4734.
28. Ortega, A., Frossard, P., Kovačević, J., Moura, J. M., & Vandergheynst, P. (2018). Graph signal processing: Overview, challenges, and applications. Proceedings of the IEEE, 106(5), 808–828.
29. Akherfi, K., Gerndt, M., & Harroud, H. (2018). Mobile cloud computing for computation offloading: Issues and challenges. Applied computing and informatics, 14(1), 1–16.
30. Rehmani, M. H., Reisslein, M., Rachedi, A., Erol-Kantarci, M., & Radenkovic, M. (2018). Integrating renewable energy resources into the smart grid: Recent developments in information and communication technologies. IEEE Transactions on Industrial Informatics, 14(7), 2814–2825.
31. Narciso, D. A., & Martins, F. G. (2020). Application of machine learning tools for energy efficiency in industry: A review. Energy Reports, 6, 1181–1199.
32. Bibri, S. E., & Krogstie, J. (2020). Environmentally data-driven smart sustainable cities: Applied innovative solutions for energy efficiency, pollution reduction, and urban metabolism. Energy Informatics, 3(1), 29.
33. Ciupageanu, D. A., Barelli, L., & Lazaroiu, G. (2020). Real-time stochastic power management strategies in hybrid renewable energy systems: A review of key applications and perspectives. Electric Power Systems Research, 187, 106497.
34. Watson, R. T., Boudreau, M. C., & Chen, A. J. (2010). Information systems and environmentally sustainable development: energy informatics and new directions for the IS community. MIS quarterly, 23–38.
35. Anand, C. K., & Amor, B. (2017). Recent developments, future challenges and new research directions in LCA of buildings: A critical review. Renewable and sustainable energy reviews, 67, 408–416.
36. Ketter, W., Collins, J., Saar-Tsechansky, M., & Marom, O. (2018). Information systems for a smart electricity grid: Emerging challenges and opportunities. ACM Transactions on Management Information Systems (TMIS), 9(3), 1–22.
37. Andreopoulou, Z. S. (2012). Green Informatics: ICT for green and Sustainability. Journal of Agricultural Informatics, 3(2).

38. Zhou, K., & Yang, S. (2016). Understanding household energy consumption behavior: The contribution of energy big data analytics. Renewable and Sustainable Energy Reviews, 56, 810–819.
39. Friedewald, M., & Raabe, O. (2011). Ubiquitous computing: An overview of technology impacts. Telematics and Informatics, 28(2), 55–65.
40. Palensky, P., & Dietrich, D. (2011). Demand side management: Demand response, intelligent energy systems, and smart loads. IEEE transactions on industrial informatics, 7(3), 381–388.
41. 46. Jiang, H., Wang, K., Wang, Y., Gao, M., & Zhang, Y. (2016). Energy big data: A survey. IEEE Access, 4, 3844–3861.
42. 45. Homeyer, N., & Gohlke, H. (2012). Free energy calculations by the molecular mechanics Poisson− Boltzmann surface area method. Molecular informatics, 31(2), 114–122.
43. 47. Usman, A., & Shami, S. H. (2013). Evolution of communication technologies for smart grid applications. Renewable and Sustainable Energy Reviews, 19, 191–199.

AI Strategies for Handling Disciplinary and Cyber Bullying in Schools

Manjit Singh Sidhu and Kirandeep Kaur Sidhu

Abstract In this chapter, we delve into the integration of artificial intelligence (AI) strategies for effectively managing disciplinary and bullying cases in schools. We begin by outlining preventative measures, harnessing AI predictive analytics to identify at-risk students and deploying virtual simulations to cultivate positive social behaviours. Utilizing natural language processing (NLP) algorithms and computer vision technologies, we explore AI role in early detection of disciplinary infractions and bullying behaviours, enabling swift intervention. Additionally, AI-driven chatbots offer immediate support and guidance to victims and perpetrators, while real-time sentiment analysis aids educators in gauging emotional climates within classrooms. Furthermore, personalized learning platforms and peer support networks leverage AI to provide ongoing support and rehabilitation for all parties involved. We conclude by emphasizing the importance of ethical considerations in deploying AI, highlighting the collaborative efforts needed to maximize its benefits in creating safe and inclusive learning environments.

Keywords Machine learning · Artificial intelligence · Disciplinary · Cyberbullying

1 Introduction

Bullying and disciplinary occurrences in schools have increased recently, especially in Malaysia. Last three years of data from the ministry shows that as of October 2023, there were 4,994 instances, 3,887 cases in 2022, and 326 cases in 2021. Parents are now worried about their children's safety and wellbeing when sending them

M. S. Sidhu (✉)
Department of Graphics and Multimedia, College of Computing and Informatics, Universiti Tenaga Nasional (UNITEN), Kajang, Selangor, Malaysia
e-mail: manjit@uniten.edu.my

K. K. Sidhu
MK-Faculty of Medicine and Health Sciences, Universiti Tunku Abdul Rahman (UTAR), Sg. Long, Kajang, Selangor, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_12

to school due to this worrying trend. Even after trying a number of approaches to address this problem, it hasn't decreased. A few factors contributing to this unfortunate state of affairs are the lack of certified counselors or psychologists in schools. Bullying incidences in schools are frequently exacerbated by the inclination to keep such incidents under wraps. It is depressing to admit that some people, whether they be classmates or teachers, take the position of not paying attention, as though the upsetting things that are happening in front of them does not matter. In recent years, cyberbullying has emerged as a significant threat to the psychosocial wellbeing of adolescents [1, 2]. Defined as intentional harmful behavior directed at a victim through electronic devices [3], cyberbullying differs markedly from traditional bullying due to its unique characteristics [4]. Unlike the repetitive nature inherent in traditional bullying dynamics, cyberbullying can occur with a single instance of posting content, which can have a lasting impact [5, 6]. Theoretical frameworks such as the "General Strain Theory (GST)", "Routine Activities Theory (RAT)" and "Social-Ecological Theory" have influenced psychosocial research on cyberbullying [7–11]. Williard [12] concluded that cyberbullying encompasses a wide range of online behaviours. Flaming, impersonation, harassment, exclusion, denigration, cyber stalking, outing/trickery and revenge porn are different classifications that categorize cyberbullying behaviors [13–16]. Within this extensive categorization, colleagues distinguish between two primary forms of cyberbullying: verbal offenses in written form and visual cyberbullying, which adolescents perceive as more harmful due to the non-consensual sharing of denigrating videos or pictures [17].

Recent research highlights the significant threat cyberbullying poses to adolescents' psychosocial wellbeing, often resulting in more severe stress, depression, and anxiety than traditional bullying [18, 19]. Studies have found a correlation between adolescents' online activities, particularly on social networking sites and gaming platforms, and an increased risk of cyberbullying [20–24]. This underscores the need for identifying protective factors and developing tailored prevention programs at primary, secondary, and tertiary levels [25]. The use of artificial intelligence (AI) and machine learning (ML) is seen as a promising approach to enhance these strategies, helping to manage and prevent bullying both in schools and online. This approach shows potential in early intervention, creating safer and more supportive environments for adolescents.

Cyberbullying has become increasingly prevalent with the widespread use of digital communication platforms. Its impact on adolescents can be severe, leading to mental health issues such as anxiety, depression, and even suicidal thoughts. Early detection of cyberbullying incidents is crucial in mitigating these negative effects and providing timely support to victims.

2 Background

Despite numerous efforts to address bullying and cyberbullying, effective prevention remains elusive, leading to persistently high rates. Researchers at MIT Media Lab, led by Karthik Dinakar, developed an algorithm using clustering and classification to detect and categorize words in online interactions [26, 27]. While this algorithm accurately categorizes content, it cannot combat cyberbullying in real-time. Various models have been developed, including a socio-linguistic model using a logical probabilistic approach to detect cyberbullying in conversational contexts [28], and Natural Language Processing (NLP) methods for identifying bullying in chat conversations [29]. Raisi's research utilizes co-trained ensembles with recurrent neural networks (RNN) and node2vec learners to detect harassment and bullying from text-based data [30]. Many studies have employed machine and deep learning techniques like k-Nearest Neighbor (kNN), Linear Regression (LR), Random Forests (RF), and others to detect harmful language. However, real-time detection remains a significant challenge.

3 Contributions

1. **Highlighting the Threats:** This research emphasizes the significant threats of bullying and cyberbullying in schools, detailing their harmful effects on students' wellbeing, academic performance, and the overall school climate. It underscores the urgent need for effective strategies to mitigate these impacts.
2. **Proposing Solutions:** The study suggests innovative solutions for addressing bullying and disciplinary issues in schools. It advocates for using AI predictive analytics to identify at-risk students early and recommends virtual simulations to foster positive social behaviors, reducing the likelihood of bullying. These strategies aim to prevent incidents through early intervention.
3. **Integrating AI in Cyberbullying:** AI is identified as crucial in managing cyberbullying by providing advanced tools for detection, intervention, and support. This includes early detection using NLP algorithms, swift intervention through computer vision, AI-driven chatbots for guidance, and real-time sentiment analysis for monitoring emotional climates.

4 AI and Cyberbullying

Eighty percent of juvenile users of the internet are engaged on social networking sites [31], increasing their exposure to harmful online experiences [32]. Despite various interventions to prevent cyberbullying [33], secondary prevention programs face challenges, as victims and bystanders often do not report incidents [34]. Machine Learning (ML) algorithms show promise for early detection and intervention, using techniques like sentiment analysis, textual and semantic analysis, and analysis of user characteristics such as gender [35–38]. These methods have been effective in identifying cyberbullying roles, involvement, and the severity of consequences [39, 40]. However, their potential remains underutilized by social scientists and educational institutions. This research seeks to highlight the benefits of ML in cyberbullying prevention, aiming to bridge the gap between the fields of computing and the behavioral sciences.

5 Objectives and Research Questions

This study aims to gather and analyze research on the application of Machine Learning (ML) for cyberbullying detection through a comprehensive review of existing literature. By focusing on previous reviews, this approach seeks to illuminate current debates and theoretical perspectives on the subject, addressing the following research questions:

1. Which features are most utilized for automated cyberbullying detection?
2. What Machine Learning techniques are employed, and the way they are assessed?
3. What effect does ML have on combatting cyberbullying?

Through this investigation, this study aims to provide insights into the state-of-the-art in ML-based cyberbullying detection, offering a critical analysis of methodologies and identifying areas for future research and development.

6 Method

This review employed a systematic search strategy and selection process. In Jan 2024, a systematic search in 3 scientific databases (namely Scopus, PsycInfo and PsycArticles) was conducted. A filter of keywords and Boolean operators was employed across databases, using terms such as “cyberbullying” (OR online bullying, OR cybervictimization), AND “machine learning” (OR artificial intelligence), AND “adolescen*” (OR teen*), AND “review”. Duplicate records were identified and removed, only English language records were included, resulting in 142 records after abstract screening. A total of 142 records were gathered from Scopus (n = 87), PsycInfo (n = 40), and PsycArticles (n = 15).

Subsequently, 109 records were excluded based on selection process of publication type (15 chapters, 3 commentary, 19 conference papers, 11 books and e-books, 43 empirical studies, and 18 dissertations). The remaining 33 records underwent full-text screening, applying a criterion focused on the core topics of cyberbullying, machine learning, and adolescence. Following this screening, 16 records were selected for further evaluation, where two independent raters conducted quality assessments and data extraction. A specialized tool was employed for quality assessment, tailored to accommodate the interdisciplinary nature of the sources. Following independent data extraction by two raters, a total of 14 review papers were deemed eligible for inclusion in the current systematic review.

7 How Teachers Handle Bullying Cases

Victim Reported Experiences: The effectiveness of bullying interventions is influenced by factors such as the intervention method used by teachers or counselors and its suitability for the specific case [41]. Some bullying cases are particularly difficult to address, especially when the bullying is so severe that the victim struggles to cope, even with support from teachers. Group bullying intensifies the victim's vulnerability, as group members can coordinate attacks collectively or individually across different locations. Effective school interventions often require collaboration with multiple perpetrators and groups whose behavior needs to be modified. Understanding how bullying severity and group involvement impact victims is crucial for addressing these issues effectively.

Severity: Severity of bullying is often inferred from the nature of the perpetrator's actions, with physical assault typically seen as more severe than verbal harassment by teachers [42]. However, students often experience multiple forms of bullying, and high levels of consistency are found in reports using multi-item scales that address different types of bullying [43]. Relying solely on reported behaviors to assess severity overlooks individual differences in students' resilience to stress. Therefore, the best way to assess bullying severity is through subjective reports from victims about the emotional impact they have experienced [44].

Bullying by Groups: Distinguishing between individual and group bullying through direct observation is challenging because it requires understanding the influence of group dynamics versus individual decisions. However, it is possible to assess whether victims perceive their bullying as driven by individual actions or group pressure. Studies using interviews or questionnaires with schoolchildren in the USA [45], Sweden [46], and Greece [47] found that students often report being bullied by individuals rather than groups. In contrast, a study in Australia based on retrospective accounts from adults suggests that victims are more likely to report being bullied by groups during their childhood [48]. Since this study focuses on students' current or recent experiences, it is expected that reports of bullying by individuals will be more common.

As adolescents increasingly spend time online, cyberbullying has become a prevalent and rising issue. Cyberbullying includes harmful actions such as hurtful social media posts, mean statements during gaming, and the creation of hate accounts to embarrass, threaten, or abuse. This type of bullying deeply infiltrates adolescents' personal lives without clear boundaries. From 2018 to 2022 (Fig. 1), cyberbullying cases have alarmingly increased, with the incidence rising from 7 to 9% among girls and from 11 to 14% among boys. Spain reports the highest prevalence of cyberbullying victimization at 57.5%, followed by Malaysia at 52.2%, Israel at 45%, and

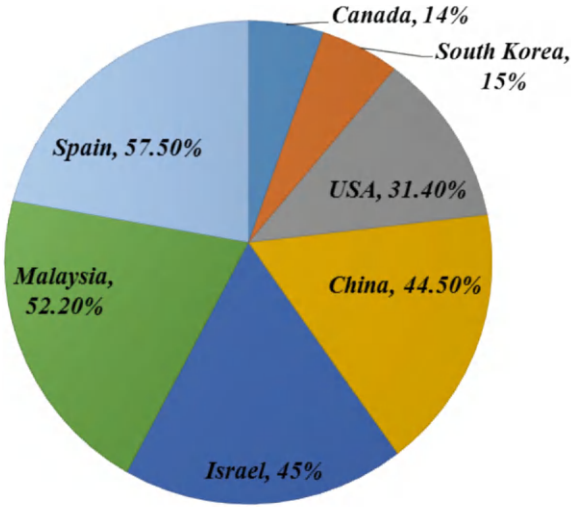


Fig. 1 Alarming increase in cyberbullying cases

China [49]. Middle and high school students, who spend considerable time online, are particularly vulnerable. Common behaviors include offensive name-calling, sharing images without consent, physical threats, spreading false rumors, and cyber-flashing. The impact on victims can be severe, leading to trauma, depression, and even suicide, as seen in tragic cases like Ryan Halligan's and others.

7.1 Cyberbullying Versus Physical Bullying

Both physical and cyberbullying are types of harassment that can have a significant impact on a person. Cyberbullying is when someone uses a digital platform, like social media, messaging applications, or gaming communities, to harass, threaten, or degrade another person while being anonymous or remote. Bullying of this kind can be widespread; it frequently crosses physical boundaries and affects victims even in the comfort of their own homes. Conversely, physical bullying refers to overt, visible acts of physical intimidation such as pushing, hitting, or other similar behaviours. While bullying in all its manifestations can result in long-term psychological problems such as anxiety and emotional dis-comfort, cyberbullying can be especially pernicious because of its ability to spread harmful content widely and its persistent nature. For victims, cyberbully-ing is a persistent problem that is sometimes made worse by the difficulty in re-moving the offensive content from the internet or managing to escape it. To stop and deal with bullying in all its forms, strong measures are needed, such as edu-cation, support networks, and stringent regulations.

Putting policies in place to stop cyberbullying in schools comes with a number of real-world obstacles. Since monitoring online interactions necessitates access to potentially sensitive personal information, data privacy is a major concern. To safeguard the rights of students, it is essential to make sure that this data is managed appropriately, in accordance with ethical standards and privacy regulations. The requirement for sizable, labelled datasets for the efficient training of ma-chine learning models presents another difficulty. These algorithms have the ability to detect bullying behaviour, but gathering and annotating these datasets requires a lot of resources and has privacy concerns.

Furthermore, it is difficult to integrate anti-bullying policies with the current ed-ucational framework. Schools frequently have established procedures and technological infrastructure, so implementing new systems calls for careful planning, manpower, and training. Effective implementation requires striking a balance between the advantages of novel interventions and the real-world constraints of educational settings.

8 Cyberbullying Detection and Prevention

Developing ML and AI-based frameworks for cyberbullying detection and prevention involves several key steps tailored to text, image, and audio data. Data collection is the first step, where sources like social media platforms, forums, and online communities are utilized. Techniques such as APIs, web scraping, and collaborations with platforms are employed, ensuring ethical considerations like data privacy and anonymity are maintained. Table 1 summarize studies found in the literature that employed various approaches and datasets in analyzing cyberbullying. The next step is preprocessing and feature extraction. For text data, this includes tokenization, stop-word removal, and using techniques like term frequency-inverse document frequency (TF-IDF) or word embedding for feature extraction. For images, pre-processing involves resizing, normalization, and using convolutional neural network (CNNs) for feature extraction. In audio data, noise reduction, segmentation, and extracting features like mel-frequency cepstral coefficients (MFCCs) and spectrograms are key steps.

Model selection and training for cyberbullying detection involve choosing suitable models based on data type. Text-based detection uses models like support vector machines (SVMs), recurrent neural networks (RNNs), or Transformer-based models like BERT. Image-based detection often employs convolutional neural networks (CNNs) or transfer learning models like ResNet. For audio data, RNNs, LSTMs, and CNNs are common. These models are trained with supervised learning, using cross-validation and hyperparameter tuning for optimization. Detection occurs in real-time, classifying content as bullying or non-bullying, followed by actions like notifying moderators and removing harmful content. Continuous improvement is essential, using metrics like accuracy, precision, recall, and F1-score to evaluate performance. Models are regularly retrained with new data, incorporating user feedback, and exploring advanced techniques to maintain effectiveness. Regular ethical reviews ensure user privacy and data protection, as summarized in Table 2.

Cyberbullying is analyzed through various theoretical perspectives. General strain theory posits that individuals may engage in bullying due to negative emotions or

Table 1 Summary of studies that employed various approaches and datasets in analyzing cyberbullying

Type of cyber bullying	Approach	Purpose	Findings/Limitation	Dataset	Refs.
Image	Pre trained transfer learning model VGG16 and InceptoionV3	Detection of image-based social cyber bullying post by extracting the hidden contextual features and generates a notification to the sender and receiver to check it and remove it	VGG16 and InceptoionV3 perform better than two dimensional deep learning based CNN model Limitation: demands an enormous amount of labelled data for training	Size: 3000 images Non Bully Class(C0): 1542 samples Bully Class (C1): 1458 samples	[50]
Text	ML, Stanford Core NLP	To let Parents to determine the text based post (tweets) on social platforms (twitter) using GUI and categorize them as bullying or non-bullying for their children To develop an app for android device	Set of 11 features were identified, found more suitable for analysis of text post Multiple Machine learning algorithms were compared and found that random forest could provide an accuracy of 92% for tweets analysis	Tweets collected 1200 Bullying tweets 200 Non bully tweets 1000	[51]
Text, audio and video	Deep learning, machine learning classifiers SVM, logistic regression, random forest	To curb cyber bullying at diverse online platforms and overcome the challenges of existing solutions	Larger models like VGG16 and 19 don't yield desired accuracy. However, Inception-V3 attains greater accuracy but require more time and resources	Images 2300 from crowd-flower website	[52]

(continued)

Table 1 (continued)

Type of cyber bullying	Approach	Purpose	Findings/Limitation	Dataset	Refs.
Text, video	Leverages cutting-edge computer vision and AI techniques, ML classification techniques	Juridical and psychological models with cutting edge technology to detect and prevent bullying Enables alerts to relevant authorities or moderators, allowing for rapid interventions and potential harm mitigation	(1) crowd analysis for potential bullying incidents using video surveillance, (2) text analysis of social media accounts for possible verbal attacks, (3) keystroke dynamics analysis to gauge potential victims' emotional states, and (4) deep fake detection to counter the spread of manipulated harmful content	Motion emotion Dataset for physical violence detection videos 32 and frames 44,000	[53]
Text	Game theory	An intervention strategy is built here to mitigate the impact of cyber-bullying	It can be used at run time ad could neutralize the effect of cyber-bullying with anti-cyber bullying users To solve the strategic-cyber-bullying-refuting optimality, forward–backward sweep method is used in this manuscript	The Network Data Repository is used here with 10,000 nodes on distinct social platforms	[54]

(continued)

Table 1 (continued)

Type of cyber bullying	Approach	Purpose	Findings/Limitation	Dataset	Refs.
Text	Parser, NLP	To design a supporting aid for net patrol people to curb bullying in japan. In order to automatically detect cyber-bullying entries, the proposed solution maximizes the category relevance	Authors have used automatic acquisition and filtering of seed words, but found this alone won't make much difference They found that single filtering is better than double filtering	Dataset with 1490 harmful and 1508 non-harmful entries (re distributed it for several cases)	[55]
Text	Quantitative (ANCOVA and Multilevel) and qualitative (Content analysis)	To enhance the adolescents' cognitive empathy and pro-sociality in cyber-bullying	Author have used quasi experimental design to imbibe empathy among young children so that they can curb the cyber-bullying effectively	221 students of 7th and 8th class were considered for the study	[56]
Text	Data mining, Fp-tree	To detect the misbehaviour pattern of student, to detect bullying & violation at school	With high level of confidence rules were obtained for detection of violence	–	[57]
Miscellaneous	Deep learning with Feature subset selection binary coyote optimization, salp swarm algorithm	To efficiently classify the information on social media to detect the CB	It has outperformed all the existing classifiers and could deliver the results with accuracy of 99.983%	3 datasets used [34, 35]	[58]
Text	Bidirectional recurrent neural network	Author proposed a GRU based architecture for CB detection	Outperforms the existing state of the art and delivers high accuracy of 98.55% approximately for both data sets	Two Dataset with 115,864 Wikipedia comments and web toxicity data 13,590:CB 102,274 NCB	[59]

(continued)

Table 1 (continued)

Type of cyber bullying	Approach	Purpose	Findings/Limitation	Dataset	Refs.
Text	Federated learning with blockchain	Author proposed to merge complex cutting edge technologies, especially blockchain, and come up with work in two-layer model	Limitation is it can't perform well with big amount of data. Hence difficult to implement	Data collected from web scrapping API 48,000 data points	[60]
Text	Random forest based CNN	Author timely detects the bullying episodes using hybrid model to aid victims with in time	Proposed model could achieve 96% and Respond in 3.4 s	Two data set from different platforms are used https://github.com/tdavidson/hate-speech-and-offensive-language	[61]
Text	Stack ensemble learning approach (Deep Learning Models)	An intervention strategy is built here to mitigate the impact of cyber-bullying	Strategic-cyber-bullying-refuting optimality, forward–backward	Ist dataset: Twitter dataset 37,373 tweets IInd dataset: Twitter and Facebook, 20,000 sentiments	[62]
Text	ML-Multinomial Naïve Bayes (MNB), Linear-SVM	AI enabled anti-CB system	Authors found that with their proposed AI based chat pod solution, bullying message will not be delivered to the target user	Dataset from multiple media-platforms like YouTube, Kaggle, Wikipedia Talk pages and Twitter Accuracy achieved –92%	[63]

Table 2 Generalized ML and AI based framework for cyberbullying detection and prevention

Text based cyber bullying detection or prevention post determination utilizing natural language processing	Image based cyber bullying detection or prevention post determination utilizing computer vision	Audio based cyber bullying detection or prevention post determination
1. Data-Collection and Pre-processing • Collect text from online platforms consisting cyberbullying comments • Remove noise/stopwords/ lemmatizing words *2. Feature Engineering* Utilizing GloVe or Word2Vec transform text into Numerical vector *3. Model Selection and Training* • SVM/Random Forest/Naïve Bayes/LSTM/CNN to determine nuances in cyber-textual data *4 Cyberbullying Detection* • Set the threshold value of the classified text based on the confidence score of the model and flag the identified bullied text	*1. Data-Collection* • From different online platforms datasets comprising images and videos must be collected *2. Pre-processing and Feature Extraction* • To make the image pertinent for cyberbullying recognition, image features need to be transformed using techniques such as CNN. Besides this labelling is also done in this step *3. Model Selection and Training* • To classify images or video frames as bullying, ML models are trained. Beside this to identify specifically violence related information object identifying models such as YOLO/SSD can be used *4. Cyberbullying Detection* • Set the threshold value of the classified text based on the confidence score of the model & flag the identified bullied text *5. Evaluation and Improvement* Using metrics such as F1-score etc. are evaluated and improved continuously	*1. Data Collection* • Audio conversations having bullying data needs to be collected from miscellaneous platforms *2. Feature Extraction* • ML techniques such as Mel-frequency Cepstral coefficients are usually used to determine acoustic features such as pitch, temporal patterns, etc. *3. Model Selection and Training* • Out of multiple models such SVM, RNNs, CNN can be used for classifying the audio as bullied or non-bullied. After this cross validate the labelled data *4. Threshold Setting* • In order to avoid the false identification of audio segment, threshold for selected model needs to be decided for better confidence score

stress. In contrast, routine activity theory emphasizes the conditions that enable bullying, such as having a vulnerable target and lacking effective supervision.

9 Technology Integration to Overcome Bullying Before It Happens Offline or Online

Integrating technology to prevent bullying (Fig. 2), both offline and online, involves a multifaceted approach that combines predictive analytics, real-time monitoring, educational tools, and community engagement. Here is an overview of how technology can be leveraged to prevent bullying before it occurs. Predictive analytics can identify potential bullying incidents before they happen by analyzing patterns and trends in behavior. Machine learning models can be trained on historical data to recognize the signs of escalating conflict or negative interactions. For instance, analyzing changes in communication patterns, such as increased frequency of negative language or sudden isolation of an individual, can help predict bullying behavior. Schools and online platforms can use these insights to intervene proactively, offering support to at-risk individuals and addressing potential bullies.

Real-time monitoring tools can detect and flag harmful behavior as it unfolds. In online environments, natural language processing (NLP) algorithms can monitor chats, social media posts, and comments for bullying-related keywords and phrases. Advanced systems can also analyze the sentiment and context of the communication to better understand the intent behind the words. For offline scenarios, schools can use surveillance cameras with AI capabilities to detect physical altercations or threatening behavior. These systems can alert authorities immediately, allowing for quick intervention. Education is a critical component of bullying prevention. Technology can facilitate educational programs that teach students, parents, and educators about the signs of bullying and effective ways to address it. Interactive apps and online

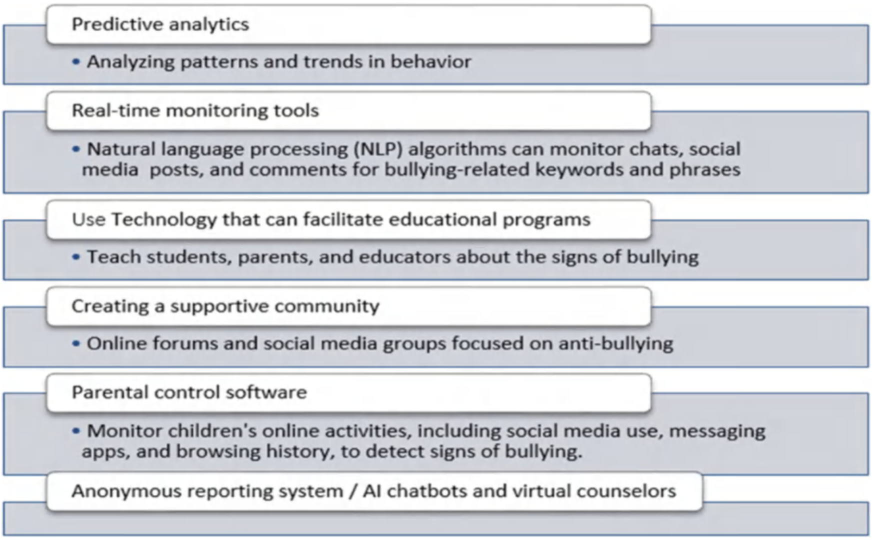


Fig. 2 Technology integration to overcome bully

platforms can offer engaging lessons on empathy, conflict resolution, and digital citizenship.

Virtual reality (VR) can provide immersive experiences, allowing individuals to understand the impact of bullying from different perspectives. These tools can help build a more informed and empathetic community that actively works to prevent bullying. Creating a supportive community is essential for preventing bullying. Technology can connect individuals through platforms designed for sharing experiences and providing mutual support. Online forums and social media groups focused on anti-bullying can offer safe spaces for victims to seek help and for others to offer advice and support. Schools and organizations can use these platforms to share resources, report incidents anonymously, and organize awareness campaigns. Community engagement platforms can foster a sense of belonging and collective responsibility, making it harder for bullying to thrive.

For younger individuals, parental control software can help prevent exposure to harmful content and interactions online. These tools can monitor children's online activities, including social media use, messaging apps, and browsing history, to detect signs of bullying. Alerts can be sent to parents or guardians when concerning behavior is identified, allowing them to take appropriate action. Additionally, these tools can offer features like screen time management and content filtering to create a safer online environment for children. Anonymous reporting systems can empower bystanders and victims to report bullying incidents without fear of retaliation. Schools and online platforms can integrate these systems into their existing infrastructure, allowing users to submit reports through apps, websites, or hotlines. This anonymity can encourage more individuals to come forward, providing authorities with the information needed to intervene early and effectively.

AI chatbots and virtual counselors can provide immediate support to individuals experiencing bullying. These systems can offer guidance on how to handle bullying situations, provide emotional support, and connect users with professional help if needed. By being available 24/7, AI-powered counselors can ensure that support is always accessible, helping to mitigate the effects of bullying and prevent it from escalating. Therefore, AI-driven chatbots can be integrated into school systems to offer immediate support to students experiencing bullying. These chatbots can gauge emotions through sentiment analysis and provide resources or escalate the issue to human counselors if necessary. Additionally, personalized learning platforms powered by AI can educate students about cyberbullying, promoting awareness and prevention strategies.

By integrating these technologies, we can create a proactive and comprehensive approach to bullying prevention. Through predictive analytics, real-time monitoring, education, community engagement, parental controls, anonymous reporting, and AI-powered support, technology can play a pivotal role in stopping bullying before it starts, both online and offline.

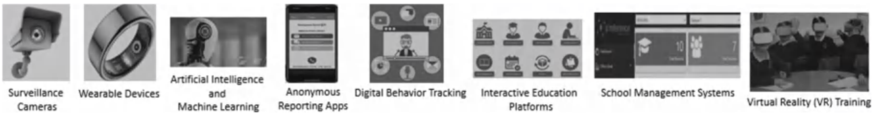


Fig. 3 Technologies to curb physical cyberbullying at school

10 Technologies to Curb Physical Cyberbullying at School

The alarming rise in physical bullying incidents within school environments has also necessitated the exploration and implementation of advanced technological solutions. Schools are increasingly turning to innovative technologies to enhance student safety and create a more secure learning atmosphere. These technologies not only deter bullying behavior but also facilitate timely intervention and support for affected students. By leveraging tools such as surveillance cameras, wearable devices, anonymous reporting apps, and artificial intelligence, schools can proactively address and mitigate physical bullying. The following sections outline various technologies and their potential to significantly curb physical bullying in educational settings. Some useful technologies are shown in Fig. 3 and explained in Table 3.

The integration of these technologies in schools can significantly mitigate physical bullying by providing real-time monitoring, enabling prompt interventions, and fostering a safer and more supportive school environment. It is essential to combine technological solutions with comprehensive policies, training, and a strong commitment from the entire school community to effectively curb physical bullying.

11 Technologies to Curb Cyberbullying at the Adolescent Age

The importance of technologies to curb cyberbullying at the adolescent age cannot be overstated. Adolescence is a critical period in a young person's development, marked by significant physical, emotional, and social changes. During this time, the negative impacts of cyberbullying can be particularly severe, affecting mental health, academic performance, and overall well-being. One of the primary reasons technology is crucial in combating cyberbullying is the pervasive nature of digital interactions among adolescents. With the widespread use of smartphones, social media platforms, and online communication tools, opportunities for cyberbullying have increased dramatically. Adolescents are more connected than ever before, and unfortunately, this connectivity can facilitate bullying behavior that is relentless, pervasive, and often anonymous. Traditional anti-bullying measures, such as adult supervision and school policies, are not sufficient to address the unique challenges posed by cyberbullying in the digital age.

Table 3 Technology integration deployment, functionality and benefits

Technologies	Deployment	Functionality	Benefits
Surveillance cameras	Strategically placed in common areas such as hallways, playgrounds, and cafeterias	Continuous monitoring and recording of activities to deter bullying incidents and provide evidence when needed	Acts as a deterrent, helps in the immediate identification of bullying incidents, and provides documentation for further action
Wearable devices	Students can wear devices like smartwatches equipped with emergency alert systems	Allows students to send distress signals to designated school authorities when they feel threatened or are in danger	Immediate response to incidents, enhanced student safety, and empowerment of students to take control of their safety
Anonymous reporting apps	Mobile applications available to students, parents, and teachers	Enables the anonymous reporting of bullying incidents, including physical bullying	Encourages reporting without fear of retaliation, provides a safe and private way to report incidents, and allows for timely intervention
Artificial intelligence and machine learning	Platforms that monitor and analyze students' behavior and interactions within the school premises	Tracks incidents of physical altercations, logs behavioral patterns, and provides reports to school counselors and administrators	Early identification of at-risk students, targeted interventions, and improved behavioral management
Interactive education platforms	Online and interactive platforms used in classrooms	Provides modules on bullying awareness, empathy training, and conflict resolution	Educates students about the impact of bullying, promotes a positive school culture, and empowers students with skills to handle conflicts non-violently
School management systems	Integrated into the school's administrative framework	Centralizes data on bullying incidents, tracks interventions, and monitors outcomes	Streamlined reporting and response process, accountability, and effectiveness tracking of anti-bullying measures
Virtual Reality (VR) training	Used in training programs for teachers and staff	Simulates bullying scenarios to train staff on recognizing and intervening in bullying situations	Enhanced preparedness, better understanding of bullying dynamics, and improved response strategies

Moreover, the effects of cyberbullying can be more insidious and long-lasting than physical bullying. Cyberbullying can occur at any time and in any place, extending beyond the school environment into the home, which should be a safe space for adolescents. The psychological impact, including anxiety, depression, and even suicidal ideation, can be profound and enduring. Technologies designed to detect and prevent cyberbullying play a crucial role in identifying harmful behavior early and providing timely intervention, thereby mitigating these adverse effects. Additionally, technologies that monitor and analyze online behavior offer valuable insights that can be used for proactive measures. Predictive analysis tools can identify patterns and trends in cyberbullying, enabling schools and parents to intervene before incidents escalate. This proactive approach not only prevents bullying but also helps create a safer and more supportive online environment for adolescents.

Another critical aspect is the empowerment and education of both students and parents. Technologies such as digital citizenship education platforms and parental control software provide essential tools and knowledge to navigate the digital world safely. These resources teach adolescents about responsible online behavior, the consequences of cyberbullying, and how to protect themselves, fostering a culture of respect and empathy in online interactions. Below are some techniques to curb cyberbullying at adolescent age.

Detection Techniques

Social media monitoring tools are essential for detecting cyberbullying among adolescents. They continuously scan platforms for harmful content using technologies like keyword detection, sentiment analysis, and image recognition. This allows for real-time identification of abusive messages and documentation of bullying patterns. The integration of artificial intelligence (AI) and machine learning into messaging apps and online forums further enhances detection. AI algorithms identify patterns and language indicative of cyberbullying, flagging potential incidents for review. This approach ensures ongoing monitoring, reduces false positives, and improves detection accuracy.

Prevention Techniques

Preventing cyberbullying is crucial for ensuring a safe online environment for adolescents. Parental control software aids in this by monitoring online activity, setting usage limits, and blocking inappropriate websites, thereby reducing exposure to harmful content and encouraging healthy online habits. Additionally, digital citizenship education platforms, integrated into school curricula, teach students about respectful online behavior, the impacts of cyberbullying, and safe internet practices. By fostering a positive online culture and equipping students with essential knowledge, these educational tools help significantly decrease cyberbullying incidents.

Intervention Techniques

Effective intervention techniques are necessary to address cyberbullying once it occurs. Anonymous reporting systems, accessible through school websites or dedicated apps, allow students to report cyberbullying incidents without revealing their identities. This ensures the safety and privacy of the reporter, encouraging victims and witnesses to come forward without fear of retaliation. Timely intervention and support can then be provided to address the reported incidents. In addition to reporting systems, online counseling and support services are crucial for supporting victims of cyberbullying. These services, available through school programs or independent platforms, provide professional counseling and a safe space for victims to discuss their experiences. This emotional support, combined with mental health resources and guidance on coping strategies, helps adolescents navigate the challenges of cyberbullying.

Predictive Analysis for Proactive Measures

Predictive analysis techniques offer a proactive approach to preventing cyberbullying. Behavioral analysis software, integrated into school management systems and social media platforms, analyzes online behavior patterns to predict potential bullying incidents. By identifying at-risk students and potential future hotspots, schools can implement proactive measures to prevent bullying. Data analytics platforms further enhance predictive capabilities by aggregating data on past bullying incidents. This data is used to identify trends and inform policy and decision-making, enabling school administrators to allocate resources effectively and improve overall school safety. These predictive analysis tools not only help in preventing cyberbullying but also in creating a comprehensive understanding of the dynamics involved, allowing for more targeted and effective interventions.

The integration of advanced detection, prevention, intervention, and predictive analysis techniques is essential in curbing cyberbullying among adolescents. By leveraging these technologies, schools and parents can create a safer and more supportive online environment. Combining technological solutions with educational initiatives and support systems fosters a positive digital culture and effectively mitigates the impact of cyberbullying, ensuring that students can engage in digital spaces without fear of harassment or harm. The integration of advanced technologies to curb cyberbullying at the adolescent age is of paramount importance. These technologies address the unique challenges of cyberbullying, provide early detection and intervention, and empower adolescents with the knowledge and tools to navigate the digital world safely. By leveraging these technologies, we can protect the mental health and well-being of adolescents, ensuring they can enjoy the benefits of digital connectivity without fear of harassment or harm.

12 Conclusion

The rising incidence of disciplinary and bullying incidents in Malaysian schools, particularly the surge in cyberbullying, presents a pressing challenge. The chapter underscores the critical need for comprehensive strategies to address both traditional and cyberbullying. Despite various efforts, the effectiveness of current interventions remains limited, highlighting gaps such as the absence of trained psychologists and the tendency to conceal incidents. Cyberbullying, with its distinct characteristics from traditional bullying, poses severe psychosocial risks to adolescents, including stress, depression, and anxiety. The adoption of theoretical frameworks like General Strain Theory and Social-Ecological Theory in understanding cyberbullying emphasizes its multifaceted nature. Innovative solutions leveraging AI and machine learning (ML) offer promising avenues for early detection, swift intervention, and personalized support.

Key contributions of this research include shedding light on the detrimental effects of bullying, proposing AI-driven solutions for early intervention, and advocating for the integration of technology in handling cyberbullying. The study's comprehensive review of ML techniques for cyberbullying detection highlights the potential of predictive analytics, real-time monitoring, and personalized learning platforms in creating a safer educational environment. However, successful implementation requires overcoming challenges such as ensuring ethical considerations in data collection, bridging the gap between social sciences and computer science, and fostering collaboration among stakeholders. The proactive use of AI and ML can significantly enhance prevention strategies, providing immediate support and fostering a supportive community for adolescents. By integrating these technologies, schools can create a proactive and comprehensive approach to bullying prevention, both online and offline, ensuring a safer and more supportive environment for students.

References

1. Olweus, D. "Cyberbullying: An overrated phenomenon?" European Journal of Developmental Psychology 9, no. 5, pp. 520–538. (2012). https://doi.org/10.1080/17405629.2012.682358
2. Baldry, A.C. Farrington, D.P. Sorrentino, A. and Blaya, C. "Cyberbullying and cybervictimization" in International Perspectives on Cyberbullying. Cham: Palgrave Macmillan, pp. 3–23. (2018) https://doi.org/10.1007/978-3-319-73263-3_1.
3. Patchin, J.W. and Hinduja, S. "Bullies move beyond the schoolyard: A preliminary look at cyberbullying." Youth Violence and Juvenile Justice 4 no. 2, pp. 148–169. (2006) https://doi.org/10.1177%2F1541204006286288
4. Olweus, D. and Limber, S.P. "Some problems with cyberbullying research." Current Opinion in Psychology 19, pp. 139–143. (2018) https://doi.org/10.1016/j.copsyc.2017.04.012
5. Olweus, D. "Bullying at school". In Aggressive behavior (pp. 97–130). Boston, MA: Springer, 1994, pp. 97–130.

6. Mishna, F. Cook, C. Gadalla, T. Daciuk, J. and Solomon, S. "Cyber bullying behaviors among middle and high school students." American Journal of Orthopsychiatry 80, no 3, pp. 362-374. (2010) https://doi.org/10.1111/j.1939-0025.2010.01040.x
7. Espelage, D.L. Rao, and Craven, R.G. "Theories of cyberbullying" in Principles of cyberbullying research, London: Routledge, 2012, pp. 77–95.
8. Agnew, R. and White, H.R. "An empirical test of general strain theory." Criminology 30 no 4, 475-500. (1992) https://doi.org/10.1111/j.1745-9125.1992.tb01113.x
9. Cohen, L.E. and Felson, M. "Social change and crime rate trends: A routine activity approach." American Sociological Review 44, no. 4, (1979), pp. 588–608. https://doi.org/10.2307/2094589
10. Navarro, J. N. and Jasinski, J.L. "Going cyber: Using routine activities theory to predict cyberbullying experiences." Sociological Spectrum 32, pp. 81–94. (2012) https://doi.org/10.1080/02732173.2012.628560
11. Vyawahare, M. and Chatterjee, M. "Taxonomy of Cyberbullying Detection and Prediction Techniques in Online Social Networks". In Data Communication and Networks, Singapore: Springer, (2020), pp. 21–37. Springer, Singapore. https://doi.org/10.1007/978-981-15-0132-6_3
12. Willard, N. E. "Cyberbullying and cyberthreats: Responding to the challenge of online social aggression, threats, and distress," Research Press, 2007.
13. Smith, P.K. del Barrio, C. and Tokunaga, R.S. "Definitions of Bullying and Cyberbullying: How Useful Are the Terms?". In Principles of cyberbullying research, (2012), London: Routledge, pp. 54-68.
14. Álvarez-García, D. Barreiro-Collazo, A. and Núñez, J. C. "Cyberaggression among adolescents: Prevalence and gender differences." Comunicar 25, (2017), pp. 89–97. https://doi.org/10.3916/C50-2017-08
15. Redmond, P Lock, J. V. and Smart, V. "Developing a cyberbullying conceptual framework for educators." Technology in Society 60, (2019). https://doi.org/10.1016/j.techsoc.2019.101223
16. E. Menesini, A. Nocentini, and P. Calussi, "The measurement of cyberbullying: Dimensional structure and relative item severity and discrimination." Cyberpsychology, Behavior, and Social Networking, (2011), pp. 267–274. https://doi.org/10.1089/cyber.2010.0002
17. Hellfeldt, K. López-Romero, L. and Andershed, H. "Cyberbullying and psychological well-being in young adolescence: the potential protective mediation effects of social support from family, friends, and teachers." International Journal of Environmental Research and Public Health 17, no. 1, (2020), pp. 45. https://doi.org/10.3390/ijerph17010045
18. Wang, J. Iannotti, R. J. and Luk, J.W. "Peer victimization and academic adjustment among early adolescents: Moderation by gender and mediation by perceived classmate support." Journal of School Health 81, (2011), pp. 386–392. https://doi.org/10.1111/j.1746-1561.2011.00606.x
19. Gamito, P.S. Morais, D.G. Oliveira, J.G Brito, R. Rosa, P.J. and de Matos, M.G. "Frequency is not enough: Patterns of use associated with risk of Internet addiction in Portuguese adolescents." Computers in Human Behavior 58, (2016), pp. 471–478. https://doi.org/10.1016/j.chb.2016.01.013 20.
20. Coyne, S.M. Padilla-Walker, L.M. Holmgren, L.G. Davis, K.M., Collier, M.K. Memmott-Elison, and A.J. Hawkins, "A meta-analysis of prosocial media on prosocial behavior, aggression, and empathic concern: A multidimensional approach." Developmental Psychology 54, (2018), pp. 331–347. https://doi.org/10.1037/dev0000412
21. Barlett, C.P. DeWitt, Maronna, B. and, Johnson, K. "Social media use as a tool to facilitate or reduce cyberbullying perpetration: A review focusing on anonymous and nonanonymous social media platforms." Violence and Gender 5, (2018), pp. 147–152. https://doi.org/10.1089/vio.2017.0057
22. 22 Chang, F.C. Chiu, C.H., Miao, N.F. Chen, P.H. Lee, C.M., Huang, T.F. and Pan, Y.C. "Online gaming and risks predict cyberbullying perpetration and victimization in adolescents." International Journal of Public Health 60, (2015), pp. 257–266. https://doi.org/10.1007/s00038-014-0643-x
23. Hilvert-Bruce, Z. and Neill, J.T. "I'm just trolling: The role of normative beliefs in aggressive behavior in online gaming." Computers in Human Behavior 102, (2020), pp. 303–311. https://doi.org/10.1016/j.chb.2019.09.003

24. Ashktorab, Z. "The Continuum of Harm" taxonomy of cyberbullying mitigation and prevention. In Online Harassment, Springer: Cham, (2018), pp. 211–227. https://doi.org/10.1007/978-3-319-78583-7_9
25. IES NCES National Center for Education Statistics. https://nces.ed.gov/fastfacts/display.asp?id=719 retrieved 20/06/2024.
26. Tomkins, Sabina, Getoor, Lise, Chen, Yunfei, Zhang, Yi. A Socio-linguistic Model for Cyberbullying Detection. (2018). 53–60. https://doi.org/10.1109/ASONAM.2018.8508294
27. Qaiser, S. and Ali, R. "Text mining: use of tf-idf to examine the relevance of words to doc ments", International Journal of Computer Applications, vol. 181, no. 1, pp. 25–29, 2018.
28. Cyberbullying Detection System with Multiple Server Configurations. IEEE International Conference on Electro/Information Technology (EIT), (2018). 0090–0095. https://doi.org/10.1109/EIT.2018.8500110
29. Rahman, S. Talukder, K. H. and Mithila, S. K. "An Empirical Study to Detect Cyberbullying with TF-IDF and Machine Learning Algorithms," International Conference on Electronics, Communications and Information Technology (ICECIT), (2021), pp. 1–4, https://doi.org/10.1109/ICECIT54077.2021.9641251.
30. Paul, S. Saha S. and Hasanuzzaman, M. "Identification of cyberbullying: A deep learning based multimodal approach", (2020) Multimedia Tools and Applications, pp. 1–20.
31. Kwolek, B. "Face detection using convolutional neural networks and Gabor filters." In 2005 International Conference on Artificial Neural Networks, Berlin: Springer, (2005), pp. 551–556. https://doi.org/ https://doi.org/10.1007/11550822_8
32. Levi, G. and Hassner, T. "Age and gender classification using convolutional neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, (2015), pp. 34–42.
33. Hampton, K. N. Goulet, L.S. Raine, L.and Purcell, K. "Social networking sites and our lives." Pew Internet and American Life Project, 2012.
34. Ybarra, J.W. and Mitchell, K.J. "Online aggressor/targets, aggressors, and targets: A comparison of associated youth characteristics." Journal of Child Psychology and Psychiatry 45, no. 7, (2004), pp. 1308 1316. https://doi.org/10.1111/j.1469-7610.2004.00328.x
35. Ang, R. P. "Cyberbullying: Its prevention and intervention strategies" in Child Safety, Welfare and Wellbeing. New Delhi: Springer, (2016), pp. 25–38.
36. Li, Q. "Cyberbullying in high schools: A study of students' behaviors and beliefs about this new phenomenon." Journal of Aggression, Maltreatment Trauma 19, no. 4, (2010), pp. 372–392. https://doi.org/10.1080/10926771003788979
37. Dani, H. Li, J. and Liu, H. "Sentiment informed cyberbullying detection in social media." In 2017 Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Cham: Springer, (2017), pp. 52–67. https://doi.org/ https://doi.org/10.1007/978-3.
38. Huang, Q. Singh, V.K. and Atrey, P.K. "Cyber bullying detection using social and textual analysis." In 2014 Proceedings of the 3rd International Workshop on Socially-Aware Multimedia, (2014), pp. 3–https://doi.org/10.1145/2661126.2661133
39. Ptaszynski, M. Masui, F. Nakajima, Y. Kimura, Y. Rzepka, R. and Araki, K. "Detecting cyberbullying with morphosemantic patterns." In 2016 Proceedings of the Joint 8th International Conference on Soft Computing and Intelligent Systems and 17th International Symposium on Advanced Intelligent Systems (SCIS-ISIS), (2016), pp. 248–255.
40. Dadvar, M. Jong, F.D. Ordelman, R. and Trieschnigg, D. "Improved cyberbullying detection using gender information." In Proceedings of the Twelfth Dutch-Belgian Information Retrieval Workshop (DIR), (2012). University of Ghent.
41. Thompson, F.; Smith, P.K. The Use and Effectiveness of Anti-Bullying Strategies in Schools; Research Report, DFE-RR098; (2011) Department for Education: London, UK.
42. Bauman, S.; Del Rio, A. Pre-service teachers' response to bullying scenarios: Comparing physical, verbal, and relational bullying. J. Educ. Psychol. (2006), 98, 219.
43. Strout, T.D.; Vessey, J.A.; DiFazio, R.L.; Ludlow, L.H. The Child Adolescent Bullying Scale (CABS): Psychometric evaluation of a new measure. Res. Nurs. Health (2018), 41, 252–264.

44. Ortega, R.; Elipe, P.; Mora-Mercha, J.A.; Luisa, M.; Genta, N.L.; Brighi, G.A.; Smith, P.K.; Thompson, F.; Tippett, P.T. The Emotional Impact of Bullying and Cyberbullying on Victims: A European Cross-National Study. Aggress. Behav. (2012), 38, 342–356.
45. Veenstra, R. Siegwar, S. Bonne, J. H. DeWinter, A. F. Verhuslt, F.C. Ormel, J. The dyadic nature of bullying and victimization: Testing a Dual-perspective Theory. Child Dev. (2007), 78, 1843–1854.
46. Thornberg, R.S. Knutsen, S. Teenagers' Explanations of Bullying. Child Youth Care Forum (2011), 3, 177–192.
47. Stylianos, S.; Plexousakis, S.S.; Kourkoutas, E.T.; Chatira, K.; Nikolopoulos, D. School Bullying and Post-traumatic Stress Disorder Symptoms: The Role of Parental Bonding. Front. Public Health (2019), 7, 75.
48. The Wesley Mission. Give Kids A Chance: No One Deserves to be Left Out; Wesley Mission: (2009) Sydney, NSW, Australia.
49. Zhu, C., Huang, S., Evans, R., & Zhang, W. Cyberbullying among adolescents and children: a comprehensive review of the global situation, risk factors, and preventive measures. (2021). Frontiers in public health, 9, 634909.
50. Roy, P. K., & Mali, F. U. Cyberbullying detection using deep transfer learning. Complex & Intelligent Systems, (2022) 8(6), 5449–5467.
51. Thun, L. J., Teh, P. L., & Cheng, C. B. CyberAid: Are your children safe from cyberbullying?. Journal of King Saud University-Computer and Information Sciences, (2022). 34(7), 4099–4108.
52. Almomani, A., Nahar, K., Alauthman, M., Al-Betar, M. A., Yaseen, Q., & Gupta, B. B. Image cyberbullying detection and recognition using transfer deep machine learning. International Journal of Cognitive Computing in Engineering, (2024). 5, 14–26.
53. Orrù, G., Galli, A., Gattulli, V., Gravina, M., Micheletto, M., Marrone, S., & Sansone, C. Development of Technologies for the Detection of (Cyber) Bullying Actions: The BullyBuster Project. Information, (2023). 14(8), 430.
54. Chu, Q., Qin, Y., Yang, L. X., & Yang, X. Game-theoretic modeling and analysis of cyberbullying spreading on OSNs. Information Sciences, (2024). 659, 120067.
55. Ptaszynski, M., Masui, F., Nitta, T., Hatakeyama, S., Kimura, Y., Rzepka, R., & Araki, K. Sustainable cyberbullying detection with category-maximized relevance of harmful phrases and double-filtered automatic optimization. International Journal of Child-Computer Interaction, (2016). 8, 15–30.
56. Ferreira, P. C., Simão, A. M. V., Paiva, A., Martinho, C., Prada, R., Ferreira, A., & Santos, F. Exploring empathy in cyberbullying with serious games. Computers & Education, (2021). 166, 104155.
57. Erlinda, D. J. P., & Devegi, M. Student Identification Based on Patterns of Association For Student Misbehaviour Using Frequent Pattern Growth Algorithms. Jurnal Teknologi dan Open Source, (2023). 6(1), 142–150.
58. Neelakandan, S., Sridevi, M., Chandrasekaran, S., Murugeswari, K., Pundir, A. K. S., Sridevi, R., & Lingaiah, T. B. (2022). Deep learning approaches for cyberbullying detection and classification on social media. Computational Intelligence and Neuroscience, 2022.
59. Raj, C., Agarwal, A., Bharathy, G., Narayan, B., & Prasad, M. Cyberbullying detection: Hybrid models based on machine learning and natural language processing techniques. Electronics, (2021). 10(22), 2810.
60. Alabdali, A. M., & Mashat, A. A novel approach toward cyberbullying with intelligent recommendations using deep learning based blockchain solution. Frontiers in Medicine, (2024). 11, 1379211.
61. Ige, T., & Adewale, S. AI powered anti-cyber bullying system using machine learning algorithm of multinomial naïve Bayes and optimized linear support vector machine. arXiv preprint arXiv:2207.11897. (2022).
62. Khairy, M., Mahmoud, T. M., & Abd-El-Hafeez, T. Automatic detection of cyberbullying and abusive language in Arabic content on social networks: a survey. (2021). Procedia Computer Science, 189, 156–166

63. Lian, A. T., Costilla Reyes, A., & Hu, X. CAPTAIN: an AI-based Chatbot for cyber-bullying prevention and intervention. In International Conference on Human-Computer Interaction (pp. 98–107). (2023) Cham: Springer Nature Switzerland.

Artificial Intelligence and Industry 4.0: Catalysts for Sustainable Organizational Development

Ala'a M. Al-Momani

Abstract Despite the rapid advancement of digital technologies and their potential to promote sustainable development, there remains a significant disconnect in their integration within organizational practices, especially in emerging markets like Jordan. This underscores the need for more localized studies in these areas to elucidate how these tools can be best utilized to address sustainability issues. Consequently, this study is intended to examine the influence of new technologies such as big data analytics, deep learning, cloud computing, and Industry 4.0 on the corporate sustainability of Jordanian SMEs. An online questionnaire collected 137 valid responses from Jordanian SMEs, which were analyzed using Partial Least Squares-Structural Equation Modeling (PLS-SEM). This research found that big data, deep learning, cloud computing, and Industry 4.0 each significantly impact organizational sustainability. The study contributes to the theoretical understanding of how emerging technologies can drive sustainability in organizational contexts, particularly within Jordanian SMEs. Moreover, for practitioners, the study underscores the benefits of leveraging these technologies to enhance sustainability practices and stimulate growth.

Keywords Artificial intelligence · Deep learning · Cloud computing · Big data analytics · Industry 4.0 · Organizational sustainability

1 Introduction

In today's world, technology significantly influences daily life, transforming industries, enhancing efficiency, and promoting sustainable development [1]. Rapid technological progress has profoundly transformed global economies and societies [2]. However, global concerns about environmental sustainability, climate change, and social responsibility have prompted organizations to seek innovative solutions [3].

A. M. Al-Momani (✉)
Department of Management Information Systems, Faculty of Business, Amman Arab University, Amman, Jordan
e-mail: a.almomani@aau.edu.jo

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_13

The rapid evolution of digital technologies has opened new avenues for enhancing sustainability efforts and driving positive transformations [4]. Business models and operations can be transformed by the integration of emerging technologies, thereby resulting in more sustainable outcomes [5]. Industry 4.0 has been gaining momentum due to examinations by researchers and practitioners on how these inventions may address problems like climate change, resource depletion, and social disparities [6, 7].

Integration of emerging technologies and embracing a mindset focused on sustainability allow organizations to explore new horizons, improve their efficiency, and strive to achieve the Sustainable Development Goals (SDGs) of the United Nations [8, 9]. Essential among these changes is Industry 4.0, which brings together such technologies as cloud computing and Big Data Analytics (BDA) [10]. With this technology, real-time data collection, predictive maintenance, and optimized resource use are possible, leading to eco-friendly and efficient operations [1, 11].

The adoption of emerging technologies for sustainability offers significant advantages by enabling more efficient resource utilization [2, 12]. Emerging technologies, for instance, provide organizations with crucial insights into consumer behavior, market trends, and environmental impacts [13]. By leveraging these tools, businesses can make informed decisions, innovate sustainable products and services, and tailor their marketing strategies accordingly [14, 15]. Beyond operational enhancements, integrating these technologies fosters innovative, resilient business models that prioritize environmental and social responsibility [16]. Embracing digital sustainability allows organizations to establish new revenue streams and distinguish themselves in competitive markets [17–19]. Consequently, this enhances their competitiveness and long-term viability while contributing to broader economic, societal, and environmental well-being [20].

1.1 Knowledge Gap

The Web of Science (WoS) database has published 484 English-language articles in journals that have titles with keywords such as "Industry 4.0", "Big Data," "Artificial Intelligence," "Cloud Computing," and "Sustainability." Among these, only three articles originate from Jordan [21–23], suggesting that the country's research community in these fields is still developing. Furthermore, a bibliometric network analysis conducted using VOSviewer software (see Fig. 1) highlights the frequency of certain keywords. "Industry 4.0" tops the list with 166 mentions, followed by "sustainability" with 105 mentions. Other significant keywords include "artificial intelligence" (90 mentions), "big data" (74 mentions), "cloud computing" (20 mentions), and "machine learning" (24 mentions). Terms specifically related to sustainability, such as "environmental sustainability" (16 mentions), "sustainable development goals" (15 mentions), and "social sustainability" (10 mentions), are also present but less frequent. This pattern indicates a strong research focus on Industry 4.0 and related technologies, with a notable emphasis on incorporating sustainability into industrial

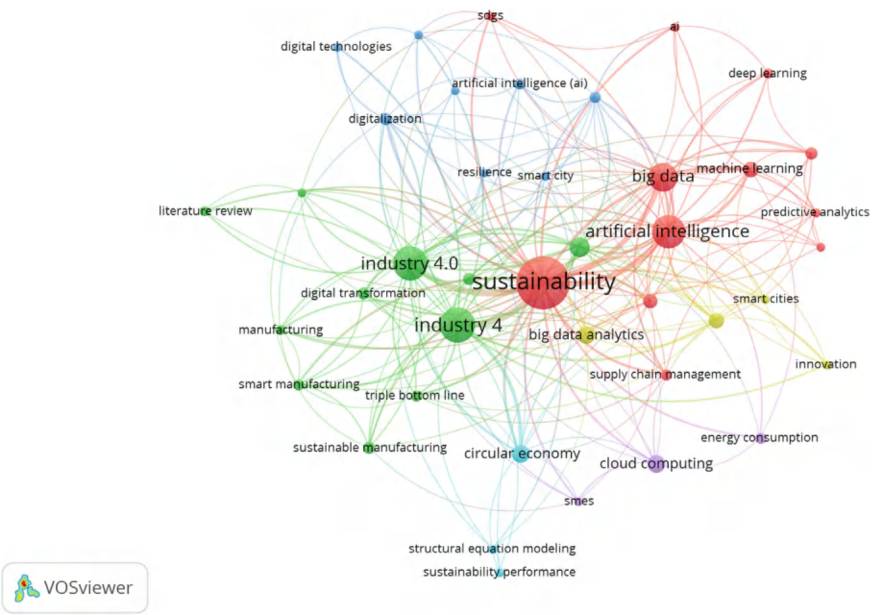


Fig. 1 The most used keywords

practices. The prevalence of these keywords underscores their importance and the growing awareness of their potential to promote sustainable development in various aspects. However, the keyword "SMEs" appears only 6 times, suggesting that, despite considerable attention to advanced technologies and sustainability, there is relatively little focus on SMEs within this research context.

1.2 Contribution and Research Questions

As businesses grapple with sustainability challenges, incorporating new technologies is a good way to bring about transforming changes and ensure that the world remains sustainable. This article examines how Jordanian Small and Medium-Sized Enterprises (SMEs) can enhance organizational sustainability using big data, deep learning, cloud computing, and Industry 4.0. Similarly, it intends to find out how improved technological advancements can enhance the sustainability of practices and encourage growth among SMEs in Jordan. It also looks into the transformative potentials these have and the practical implications for supporting sustainable development in contemporary business settings. Therefore, this study asks the following queries:

RQ1. What is the impact of big data analytics usage on organizational sustainability?

RQ2. What is the impact of deep learning technologies usage on organizational sustainability?
RQ3. What is the influence of cloud computing usage on organizational sustainability?
RQ4. What is the influence of Industry 4.0 on organizational sustainability?

2 Model and Hypotheses Development

2.1 Technology-Environmental, Economic, and Social Sustainability (T-EESST) Theory

While traditional theories emphasize factors affecting technology usage, the Technology-Environmental, Economic, and Social Sustainability (T-EESST) theory by Al-Emran [24] changes the focus to the impacts of technology application on social, economic, and environmental sustainability. This shift is vital because using technology can have both negative and positive effects [25], which existing IS models fail to consider [24]. T-EESST is needed as it becomes increasingly dependent on technology use with growing evidence of its impact on society, the economy, and the environment [24]. The T-EESST theory goes beyond mere "usage" to "impact," thereby addressing wider implications of the technology in sustainable development for a more comprehensive understanding of its role [26].

In line with this view, a comprehensive review was conducted to determine expected constructs and their interrelations in the research framework. Figure 2 shows a proposed model. This study contributes to the literature by extending the T-EESST theory with new terms such as big data, deep learning, cloud computing, and Industry 4.0.

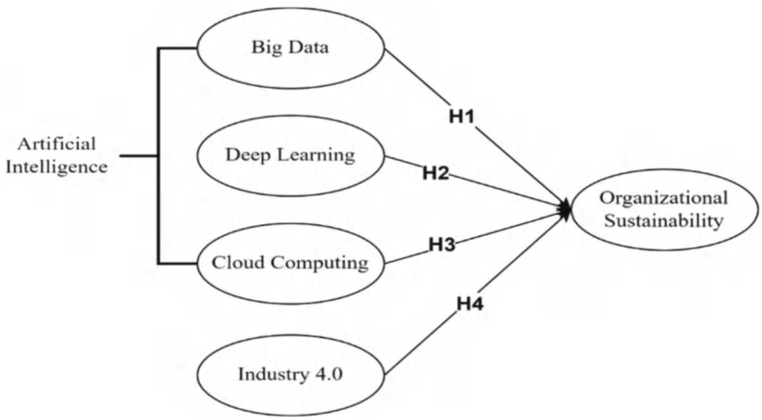


Fig. 2 The proposed model

Based on Fig. 2, the T-EESST framework incorporates elements directly linked to sustainability—including environmental, economic, and social dimensions—as outcomes from the utilization of big data, deep learning, cloud computing, and Industry 4.0 technologies. This approach aims to capture how technological practices influence sustainable development accurately. The central hypothesis of the model investigates the effects of adopting these technologies on organizational sustainability, suggesting that each contributes significantly to creating a robust and flexible business landscape, essential for the enduring prosperity of SMEs in Jordan.

2.2 *Big Data and Organizational Sustainability*

Big data is identified by its enormous size, fast speed, and wide variety of sources, which require sophisticated technologies and analytical techniques to convert them into valuable insights. This definition highlights the unique attributes of Big Data that set it apart from conventional data processing approaches, promoting a cohesive understanding and advancement of the concept within contemporary frameworks [27]. The effect of Big Data Analytics Capabilities (BDAC) on promoting organizational innovation and achieving sustainability was investigated across the hotel industry's economic, social, and environmental dimensions by Aziz et al. [28]. It found out the key factors influencing BDAC's critical role in fostering innovation and sustainability. Similarly, Ali et al. [29] highlighted how big data analytics contributed to sustainable product development, thus enhancing corporate performance. This emphasized the significance of big data technology in facilitating sustainable product development initiatives to improve organizational sustainability for future societies. Building on this, Hao et al. [30] studied the relationship between big data, analytics capability, and sustainable innovation performance. Their research revealed a dilemma between big data and analytics capability, emphasizing the need for an optimum balance. According to the research, whether or not innovation driven by big data could succeed depends on the analytics capability level, which offers guidance to organizations in their quest to enhance sustainability through strategic resource allocation decisions. Finally, Pesqueira et al. [31] examined how big data technology was transforming professionals' skill sets, promoting sustainable development by addressing key organizational challenges. Therefore, this finding supports prior studies' assertion about embracing Big Data competencies to enhance technological advancement and diffusion in those sectors. Thus, we have the following hypothesis:

H1: Big data positively influences sustainability.

2.3 Deep Learning and Organizational Sustainability

Boldrini et al. [32] describe Deep Learning (DL) as utilizing deep neural networks to construct models. According to Arrieta et al. [33], deep learning constitutes a subset of machine learning methods focused on modeling and interpreting complex data patterns using multi-layered neural networks. Parmar et al. [34] highlight deep learning's capability to extract various levels of features from data without explicit specification. Bilan et al. [35] underscored the importance of organizational learning for firm sustainability, finding that integrating deep learning and AI could enhance sustainability by promoting innovation and effective knowledge management. Mahmood [36] examined the synergy between digital leadership and AI, revealing that this integration boosted organizational performance and productivity. This quantitative research illuminated how digital transformation, AI, and digital leadership collectively drove sustainable outcomes for businesses. In addition, Wajiya et al. [37] investigated the combination of AI capabilities with organizational maturity to enhance financial sustainability in telecommunications firms. This research demonstrated how integrating AI with organizational maturity improved financial sustainability, focusing on practical applications in the real world. Accordingly, the following hypothesis is:

H2: Deep learning positively influences organizational sustainability.

2.4 Cloud Computing and Organizational Sustainability

Girs et al. [38] describe cloud computing as offering remote access to computing resources for individual users and businesses. Vaquero et al. [39] characterize cloud computing as an innovative model for provisioning computing resources, notable for its distinctive features and capabilities. Al-Mutawa and Al Mubarak [40] investigated the effects of cloud computing on sustainability among SMEs. The findings of the study showed that factors such as cutting down costs, simplicity, dependability, and collaboration greatly impacted the sustainability of SME's. Olayah [41] asserted that cloud computing enabled SMEs to foster business growth, improve performance, and achieve sustainability by reducing costs and enhancing business process efficiency. The research indicated that cloud computing positively affected organizational performance's financial, customer, and operational dimensions, with cost savings being a primary driver for sustainability among SMEs [9, 42]. Cloud computing adoption in the insurance industry was studied by Ajowi and Reuben [43] to highlight the role of technological, organizational, and environmental factors on cloud service adoption. Hence, this hypothesis is as follows:

H3: Cloud computing positively influences organizational sustainability.

2.5 Industry 4.0 and Organizational Sustainability

According to Öztemel and Gürsev [44], Industry 4.0 signifies the fourth industrial revolution, integrating digital technologies into manufacturing for automation, data exchange, and intelligent decision-making. Bonilla et al. [45], explored the implications of Industry 4.0 on sustainability, focusing on technologies like the Internet of Things and big data analytics. The study underscored the complex impacts and challenges these advancements pose, emphasizing the need to understand their influence on organizational sustainability and environmental concerns. Sharma et al. [46], described Industry 4.0 as a transformative era characterized by AI, robotics, and IoT, revolutionizing production processes and significantly contributing to sustainable development. Acioli et al. [47], investigated the application of Industry 4.0 in sustainable supply chains during the COVID-19 pandemic, highlighting its effects on supply chain processes. Their research identified opportunities, challenges, and gaps in leveraging Industry 4.0 to enhance sustainable supply chain performance, emphasizing its role in advancing organizational sustainability through technological innovations in supply chain management. Nikolova [48], examined the impact of Industry 4.0 on sustainable development, particularly focusing on challenges SMEs face in adopting Fourth Industrial Revolution processes. The study stressed the importance of addressing these challenges across industrial sectors and other relevant domains to ensure sustainable development, offering recommendations for integrating Industry 4.0 with future industrial advancements. Thus, the following hypothesis is:

H4: Industry 4.0 positively influences organizational sustainability.

3 Methodology and Data Collection

3.1 Development of Instruments

The measurement instruments for this study's constructs were specifically tailored to fit its context. Items related to big data, deep learning, cloud computing, and Industry 4.0 were adapted from Abdullah and Almaqtari [49], and organizational sustainability from Al-Emran [24]. These constructs were evaluated on a 7-point Likert scale. A pilot test was done to check the internal consistency of the questionnaire that had been availed for this purpose. It was then dispatched to 55 Jordanian SMEs which resulted in 43 usable questionnaires with all items completed. All constructs exhibited adequate reliability and validity as evidenced by each Cronbach alpha value exceeding the recommended threshold of 0.70 and factor analyses for all constructs surpassing the recommended value of 0.40.

3.2 Data Collection

Data was collected from March to May 2024 through an online survey among Jordanian SMEs across different sectors, including manufacturing, services, technology, etc. Furthermore, power analysis advised that at least 129 respondents should use SEM in such studies since they employ techniques like PLS-SEM [46]. To ensure good representation in terms of different industries and regions within Jordan, a sample size of 200 SMEs was arrived at using stratified random sampling from a larger population. The same questions were sent out to two hundred managers, who responded with one hundred sixty-two replies after excluding twenty-five incomplete ones. Eventually, due to these inadequate data, this study focused on only those responses that yielded 27 valid useable responses, representing an overall response rate of approximately sixty-eight-point five percent (68.5%).

4 Data Analysis

Data was analyzed using the partial least squares (PLS) approach, and the SmartPLS 4 software package was used to assist in performing the analysis [50, 51]. For this study, Structural equation modeling (SEM) with a two-step process involving separate measurement and structural models was used, as described by Hair et al. [52].

4.1 Assessment of Measurement Model

Based on Table 1, all indicators exhibited ($0.718 < \text{loadings} < 0.968$) and acceptable internal consistency ($759 < \alpha < 0.951$), which means that the researchers' constructs are reliable ones. Moreover, composite reliability (CR) is achieved, where all values of scores showed that it lies between ($0.847 < \text{CR} < 0.966$). Furthermore, all measurements meet the criterion of convergence validity indicated by average variance extracted (AVE) having values of ($0.649 < \text{AVE} < 704$). Furthermore, Table 2 indicates that all Heterotrait- Monotrait values do not exceed 0.85; therefore, discriminant validity is carried out as per Henseler et al. [53].

4.2 Structural Model Assessment

The structural model evaluation results are shown in Fig. 3 and Table 3. Big data, deep learning, cloud computing, and Industry 4.0 predict organizational sustainability, with all hypotheses supported. The path coefficients indicate that big data

Table 1 Loadings, Cronbach's alpha (α), internal consistency, CR and AVE

Variable	Item	Loading (>0.7)	α (>0.7)	CR (>0.7)	AVE (>0.5)
Big data	BD1	0.954	0.948	0.966	0.904
	BD2	0.930			
	BD3	0.968			
Cloud computing	CC1	0.930	0.917	0.947	0.857
	CC2	0.918			
	CC3	0.930			
Deep learning	DL1	0.779	0.759	0.847	0.649
	DL2	0.774			
	DL3	0.861			
Industry 4.0	INDS1	0.932	0.945	0.960	0.858
	INDS2	0.942			
	INDS3	0.954			
	INDS4	0.875			
Organizational sustainability	OS1	0.762	0.951	0.957	0.649
	OS2	0.801			
	OS3	0.817			
	OS4	0.718			
	OS5	0.769			
	OS6	0.799			
	OS7	0.800			
	OS8	0.776			
	OS9	0.832			
	OS10	0.856			
	OS11	0.858			
	OS12	0.869			

Table 2 The HTMT statistics showing results for testing discriminant validity

	BD	CC	DL	INDS	OS
BD					
CC	0.162				
DL	0.346	0.387			
INDS	0.101	0.056	0.137		
OS	0.117	0.600	0.713	0.115	

Fig. 3 Results of the PLS algorithm

Table 3 Results of the structural path analysis

Hypothesis	Std beta	Std error	t-value	p-value	Decision
H1	−0.083	0.048	1.724	0.042	Supported
H2	0.360	0.063	5.747	0.000	Supported
H3	0.560	0.060	9.295	0.000	Supported
H4	0.079	0.046	1.736	0.041	Supported

and Industry 4.0 have a greater impact on organizational sustainability than the other factors.

5 Discussion

This research found that big data exerts a significant impact on organizational sustainability, a hypothesis strongly corroborated by existing literature [28–31]. Thus, big data enables informed decision-making for businesses by analyzing large datasets.

This will result in more efficient utilization of resources, less waste, and a better appreciation of clients and market trends that are fundamental for sustainability. The second hypothesis, also asserting that deep learning positively impacts organizational sustainability, is supported by previous literature [34–37]. In this case, deep learning can support the creation of new products and services that are more efficient and sustainable hence providing a competitive edge to organizations on their way to sustainability. In addition, the third hypothesis that cloud computing positively impacts organizational sustainability is significantly supported by the current study, aligning with the prior literature [40, 41, 43]. This indicates that cloud computing helps reduce the need for physical infrastructure, leading to lower energy consumption and reduced carbon footprint. Moreover, organizations can scale their resources up or down as needed, ensuring they use only what they need and when they need it, which supports economic sustainability and reduces waste. Finally, the proposition that Industry 4.0 positively influences organizational sustainability is strongly supported by this study and aligns with the existing literature [45–48]. Accordingly, Industry 4.0 can significantly enhance productivity and efficiency, reduce resource use, and minimize waste. Furthermore, through smart factories and automated processes, Industry 4.0 can ensure that manufacturing processes are as efficient as possible, reducing errors, material waste, and energy consumption. The findings consistently suggest that when effectively integrated into organizational operations, big data, deep learning, cloud computing, and Industry 4.0 can substantially boost performance, thereby advancing overall sustainability within organizations.

6 Conclusion

6.1 Theoretical Implications

Incorporating emerging technologies like big data, deep learning, cloud computing, and Industry 4.0 into organizational operations highlights their potential to transform traditional business models. Al-Emran's T-EESST theory shifts the focus from simply using technology to understanding its broader social, economic, and environmental impacts. This theory is crucial for capturing both the positive and negative outcomes of technology use, offering a comprehensive framework for sustainable development. The current study contributes to this theoretical landscape by investigating how these advanced technologies can enhance organizational sustainability, especially within Jordanian SMEs.

6.2 Practical Implications

For practitioners, this study reveals how leveraging emerging technologies can bolster sustainability practices and spur organizational growth. Integrating Big Data analytics can ignite innovation and refine product development, while Deep Learning can enhance knowledge management and productivity. It is crucial for small-sized businesses because cloud computing provides cost reduction, ease of use as well as improved business process efficiency. Industry 4.0 technologies including AI and IoT can transform supply chain management as well as production processes. By doing so, organizations can significantly improve operational efficiency linked with sustainable development and aligned to SDGs.

6.3 Limitations and Future Directions

The study has limitations despite its invaluable contribution. The focus on Jordanian SMEs may mean that the findings cannot be generalized to other areas or larger firms. Therefore, future research should include various geographical locations and organizational sizes to confirm or challenge the results. Additionally, while the study highlights the positive effects of emerging technologies on sustainability, it does not thoroughly examine the potential challenges and risks associated with their adoption. Future studies should explore these aspects to provide a balanced perspective on the implications of technology integration.

References

1. R. T. Al-Majali *et al.*, "Competitive Advantage Through Analytical Capabilities: An Examination of the Relationship Between Business Analytics Capabilities and Competitiveness of Jordanian SMEs," in *Studies in Systems, Decision and Control*, Springer International Publishing, 2024, pp. 1165–1178. https://doi.org/10.1007/978-3-031-56586-1_85.
2. A. M. Al-Momani and T. Ramayah, "Adoption of Artificial Intelligence in Education: A Systematic Literature Review," in *Studies in Computational Intelligence*, M. A. Al-Sharafi, M. Al-Emran, G. W.-H. Tan, and K.-B. Ooi, Eds. Cham: Springer Nature Switzerland, 2024, pp. 117–135. https://doi.org/10.1007/978-3-031-61463-7_7.
3. I. Arpaci, K. Karataş, G. Zeybek, and A. Haktanir, "Environmental Attitude, Global Social Responsibility, and Digital Literacy: Predictors of Green Purchase Intentions Among Emerging Adults," in *Studies in Computational Intelligence*, M. A. Al-Sharafi, M. Al-Emran, G. W.-H. Tan, and K.-B. Ooi, Eds. Cham: Springer Nature Switzerland, 2024, pp. 79–96. https://doi.org/10.1007/978-3-031-61463-7_5.
4. M. A. Hajar, A. A. Alkahtani, D. N. Ibrahim, M. R. Darun, M. A. Al-Sharafi, and S. K. Tiong, "The approach of value innovation towards superior performance, competitive advantage, and sustainable growth: A systematic literature review," *Sustain.*, vol. 13, no. 18, 2021, https://doi.org/10.3390/su131810131.
5. A. Alsyouf *et al.*, "The Use of a Technology Acceptance Model (TAM) to Predict Patients' Usage of a Personal Health Record System: The Role of Security, Privacy, and Usability," *Int.*

J. Environ. Res. Public Health, vol. 20, no. 2, p. 1347, Jan. 2023, https://doi.org/10.3390/ijerph20021347.
6. C. Bai, P. Dallasega, G. Orzes, and J. Sarkis, "Industry 4.0 technologies assessment: A sustainability perspective," *Int. J. Prod. Econ.*, vol. 229, p. 107776, 2020.
7. U. Ravi and S. Chelliah, "Emerging Technology Opportunities and Challenges towards Organization and National Development," *Int. J. Bus. Technol. Manag.*, vol. 5, no. 2, pp. 104–113, 2023.
8. R. K. Abdelhamid and M. Maqableh, "The Power of Persuasion: Exploring Social Engineering in the Digital Age," in *Studies in Computational Intelligence*, M. A. Al-Sharafi, M. Al-Emran, G. W.-H. Tan, and K.-B. Ooi, Eds. Cham: Springer Nature Switzerland, 2024, pp. 307–330. https://doi.org/10.1007/978-3-031-61463-7_16.
9. M. A. Al-Sharafi, M. Iranmanesh, M. Al-Emran, A. I. Alzahrani, F. Herzallah, and N. Jamil, "Determinants of cloud computing integration and its impact on sustainable performance in SMEs: An empirical investigation using the SEM-ANN approach," *Heliyon*, vol. 9, no. 5, p. e16299, May 2023, https://doi.org/10.1016/j.heliyon.2023.e16299.
10. Z. Li, S. Rasool, M. F. Cavus, and W. Shahid, "Sustaining the future: How green capabilities and digitalization drive sustainability in modern business," *Heliyon*, vol. 10, no. 1, p. e24158, Jan. 2024, https://doi.org/10.1016/j.heliyon.2024.e24158.
11. Z. Hong and K. Xiao, "Digital economy structuring for sustainable development: the role of blockchain and artificial intelligence in improving supply chain and reducing negative environmental impacts," *Sci. Rep.*, vol. 14, no. 1, p. 3912, Feb. 2024, https://doi.org/10.1038/s41598-024-53760-3.
12. M. M. Ayyash, F. A. T. Herzallah, and M. A. Al-Sharafi, "Arab cultural dimensions model for e-government services adoption in public sector organisations: An empirical examination," *Electron. Gov.*, vol. 18, no. 1, pp. 9–44, 2022, https://doi.org/10.1504/EG.2022.119608.
13. M. Cooper, "Procurement Innovation : An Inquiry into the Adoption of Emerging Technologies in Supply Chain Management Procurement Innovation : An Inquiry into the Adoption of Emerging Technologies in Supply Chain Management," 2024, https://doi.org/10.20944/preprints202407.0384.v1.
14. A. H. Gharaibeh and N. M. Kassim, "Toward a Theoretical Framework of Business Analytics on Organizational Performance BT," in *Studies in Computational Intelligence*, M. A. Al-Sharafi, M. Al-Emran, G. W.-H. Tan, and K.-B. Ooi, Eds. Cham: Springer Nature Switzerland, 2024, pp. 351–368. https://doi.org/10.1007/978-3-031-61463-7_18.
15. Z. Tian, L. Qiu, and L. Wang, "Drivers and influencers of blockchain and cloud-based business sustainability accounting in China: Enhancing practices and promoting adoption," *PLoS One*, vol. 19, no. 1, p. e0295802, Jan. 2024, https://doi.org/10.1371/journal.pone.0295802.
16. C. Acciarini, F. Borelli, F. Capo, F. Cappa, and C. Sarrocco, "Can digitalization favour the emergence of innovative and sustainable business models? A qualitative exploration in the automotive sector," *J. Strateg. Manag.*, vol. 15, no. 3, pp. 335–352, Jul. 2022, https://doi.org/10.1108/JSMA-02-2021-0033.
17. P. H. Eng and R. L. Liu, "Will Intelligent Technologies Replace Humans in the Future? An Exploratory Study for Likelihood of Intelligent Technologies to Replace Humans in the Future," in *Studies in Computational Intelligence*, M. A. Al-Sharafi, M. Al-Emran, G. W.-H. Tan, and K.-B. Ooi, Eds. Cham: Springer Nature Switzerland, 2024, pp. 171–202. https://doi.org/10.1007/978-3-031-61463-7_9.
18. A. A. A. Haija *et al.*, "The Impact of Artificial Intelligence on Corporate Performance: A Study of Financial and Non-financial aspects in the Jordanian Service Sector," in *Studies in Computational Intelligence*, Springer International Publishing, 2024.
19. E. Mbunge, B. Muchemwa, and J. Batani, "Sensors and healthcare 5.0: transformative shift in virtual care through emerging digital health technologies," *Glob. Heal. J.*, vol. 5, no. 4, pp. 169–177, 2021.
20. R. Godina, I. Ribeiro, F. Matos, B. T. Ferreira, H. Carvalho, and P. Peças, "Impact Assessment of Additive Manufacturing on Sustainable Business Models in Industry 4.0 Context," *Sustainability*, vol. 12, no. 17, p. 7066, Aug. 2020, https://doi.org/10.3390/su12177066.

21. S. Matarneh, A. Z. Piprani, R. M. Ellahi, D. N. Nguyen, T. M. Le, and S. Nazir, "Industry 4.0 technologies and Circular Economy Synergies: Enhancing Corporate Sustainability through Sustainable Supply Chain Integration and Flexibility," *Environ. Technol. Innov.*, p. 103723, 2024.
22. K. R. Malik, Y. Sam, M. Hussain, and A. Abuarqoub, "A methodology for real-time data sustainability in smart city: Towards inferencing and analytics for big-data," *Sustain. Cities Soc.*, vol. 39, pp. 548–556, 2018.
23. S. AlZu'bi, M. Alsmirat, M. Al-Ayyoub, and Y. Jararweh, "Artificial intelligence enabling water desalination sustainability optimization," in *2019 7th international renewable and sustainable energy conference (IRSEC)*, 2019, pp. 1–4.
24. M. Al-Emran, "Beyond technology acceptance: Development and evaluation of technology-environmental, economic, and social sustainability theory," *Technol. Soc.*, p. 102383, 2023.
25. M. Al-Emran, A. A. AlQudah, G. A. Abbasi, M. A. Al-Sharafi, and M. Iranmanesh, "Determinants of Using AI-Based Chatbots for Knowledge Sharing: Evidence From PLS-SEM and Fuzzy Sets (fsQCA)," *IEEE Trans. Eng. Manag.*, pp. 1–15, 2023, https://doi.org/10.1109/tem.2023.3237789.
26. M. Al-Emran and C. Griffy-Brown, "The role of technology adoption in sustainable development: Overview, opportunities, challenges, and future research agendas," *Technol. Soc.*, vol. 73, p. 102240, 2023, https://doi.org/10.1016/J.TECHSOC.2023.102240.
27. A. De Mauro, M. Greco, and M. Grimaldi, "A Formal Definition of Big Data Based on Its Essential Features," *Libr. Rev.*, vol. 65, no. 3, pp. 122–135, 2016, https://doi.org/10.1108/lr-06-2015-0061.
28. N. A. Aziz, A. Al Mamun, M. N. H. Reza, and F. Naznen, "The Impact of Big Data Analytics On innovation Capability and Sustainability Performance Of hotels: Evidence From An emerging Economy," *J. Enterp. Inf. Manag.*, vol. 37, no. 3, pp. 1044–1068, 2024, https://doi.org/10.1108/jeim-07-2023-0354.
29. S. Ali, P. Poulová, F. Yasmin, M. Danish, M. W. Akhtar, and H. M. U. Javed, "How Big Data Analytics Boosts Organizational Performance: The Mediating Role of the Sustainable Product Development," *J. Open Innov. Technol. Mark. Complex.*, vol. 6, no. 4, p. 190, 2020, https://doi.org/10.3390/joitmc6040190.
30. S. Hao, H. Zhang, and M. Song, "Big Data, Big Data Analytics Capability, and Sustainable Innovation Performance," *Sustainability*, vol. 11, no. 24, p. 7145, 2019, https://doi.org/10.3390/su11247145.
31. A. M. Pesqueira, M. J. Sousa, and Á. Rocha, "Big Data Skills Sustainable Development in Healthcare and Pharmaceuticals," *J. Med. Syst.*, vol. 44, no. 11, 2020, https://doi.org/10.1007/s10916-020-01665-9.
32. L. Boldrini, J. Bibault, C. Masciocchi, Y. Shen, and M.-I. Bittner, "Deep Learning: A Review for the Radiation Oncologist," *Front. Oncol.*, vol. 9, 2019, https://doi.org/10.3389/fonc.2019.00977.
33. A. B. Arrieta *et al.*, "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges Toward Responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020, https://doi.org/10.1016/j.inffus.2019.12.012.
34. C. Parmar, J. D. Barry, A. Hosny, J. Quackenbush, and H. J. W. L. Aerts, "Data Analysis Strategies in Medical Imaging," *Clin. Cancer Res.*, vol. 24, no. 15, pp. 3492–3499, 2018, https://doi.org/10.1158/1078-0432.ccr-18-0385.
35. O. Isaac, Z. Abdullah, T. Ramayah, and A. M. Mutahar, " Internet Usage within Government Institutions in Yemen: An Extended Technology Acceptance Model (TAM) with Internet Self-Efficacy and Performance Impact," *Sci. Int.*, vol. 29, no. 4, pp. 737–747, 2017.
36. G. Mahmood, M. S. Khakwani, A. Zafar, and Z. Abbas, "Impact of Digital Transformation and AI Through Fostering Digital Leadership Excellence: A Focus on Sustainable Organizational Performance," *J. Account. Financ. Emerg. Econ.*, vol. 10, no. 1, 2024, https://doi.org/10.26710/jafee.v10i1.2925.

37. F. M. Wajiya and S. H. Saleh, "Integrating Artificial Intelligence Capabilities and Organizational Maturity on Enhancing Financial Sustainability - A Field Study in Iraqi Telecommunications Companies (Zain Iraq and Asia Cell)," *Int. J. Relig.*, vol. 5, no. 10, pp. 491–502, 2024, https://doi.org/10.61707/jj0bmj06.
38. S. Girs, S. Sentilles, S. A. Asadollah, M. Ashjaei, and S. Mubeen, "A Systematic Literature Study on Definition and Modeling of Service-Level Agreements for Cloud Services in IoT," *Ieee Access*, vol. 8, pp. 134498–134513, 2020, https://doi.org/10.1109/access.2020.3011483.
39. L. M. Vaquero, L. Rodero-Merino, J. Cáceres, and M. Lindner, "A Break in the Clouds," *Acm Sigcomm Comput. Commun. Rev.*, vol. 39, no. 1, pp. 50–55, 2008, https://doi.org/10.1145/1496091.1496100.
40. B. Al-Mutawa and M. M. Saeed Al Mubarak, "Impact of Cloud Computing as a Digital Technology on SMEs Sustainability," *Compet. Rev. an Int. Bus. J. Inc. J. Glob. Compet.*, vol. 34, no. 1, pp. 72–91, 2023, https://doi.org/10.1108/cr-09-2022-0142.
41. F. Olayah, "A Conceptual Modeling for Achieving a Performance and Sustainability on SMEs Using Cloud Computing Technology," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 1.4, pp. 236–242, 2019, https://doi.org/10.30534/ijatcse/2019/3681.42019.
42. M. A. Al-Sharafi, R. A. Arshah, E. A. Abu-Shanab, and Q. Alajmi, "The Effect of Sustained Use of Cloud-Based Business Services on Organizations' Performance: Evidence from SMEs in Malaysia," in *5th International Conference on Information Management, ICIM 2019*, 2019, pp. 285–291. https://doi.org/10.1109/INFOMAN.2019.8714699.
43. B. E. N. Ajowi and J. Reuben, "Technological, Organizational and Environmental Factors on the Adoption of Cloud Computing in Insurance Industries in Kenya: A Case Study of Uap-Old Mutual, Nairobi, Kenya," vol. 6, no. 4, 2019, https://doi.org/10.61426/sjbcm.v6i4.1470.
44. E. Öztemel and S. Gürsev, "Literature Review of Industry 4.0 and Related Technologies," *J. Intell. Manuf.*, vol. 31, no. 1, pp. 127–182, 2018, https://doi.org/10.1007/s10845-018-1433-8.
45. S. H. Bonilla, H. R. O. Silva, M. T. da Silva, R. F. Gonçalves, and J. B. Sacomano, "Industry 4.0 and Sustainability Implications: A Scenario-Based Analysis of the Impacts and Challenges," *Sustainability*, vol. 10, no. 10, p. 3740, 2018, https://doi.org/10.3390/su10103740.
46. D. K. Sharma, P. K. Patel, and M. Shah, "A Comprehensive Study on Industry 4.0 in the Pharmaceutical Industry for Sustainable Development," *Environ. Sci. Pollut. Res.*, vol. 30, no. 39, pp. 90088–90098, 2023, https://doi.org/10.1007/s11356-023-26856-y.
47. C. Acioli, A. Scavarda, and A. da C. Reis, "Applying Industry 4.0 Technologies in the COVID–19 Sustainable Chains," *Int. J. Product. Perform. Manag.*, vol. 70, no. 5, pp. 988–1016, 2021, https://doi.org/10.1108/ijppm-03-2020-0137.
48. N. Nikolova, "Industry 4.0 - Development and Consequences for Sustainable Development of Bulgaria," *Environ. Technol. Resour. Proc. Int. Sci. Pract. Conf.*, vol. 3, pp. 245–251, 2021, https://doi.org/10.17770/etr2021vol3.6543.
49. A. A. H. Abdullah and F. A. Almaqtari, "The impact of artificial intelligence and Industry 4.0 on transforming accounting and auditing practices," *J. Open Innov. Technol. Mark. Complex.*, vol. 10, no. 1, p. 100218, Mar. 2024, https://doi.org/10.1016/j.joitmc.2024.100218.
50. C. M. Ringle, S. Wende, and J.-M. Becker, "'SmartPLS 4.,'" 2022.
51. M. Al-Emran, V. Mezhuyev, and A. Kamaludin, *PLS-SEM in Information Systems Research: A Comprehensive Methodological Reference*, vol. 845. Springer Nature, 2019. https://doi.org/10.1007/978-3-319-99010-1_59.
52. J. F. Hair Jr, G. T. M. Hult, C. M. Ringle, and M. Sarstedt, *A primer on partial least squares structural equation modeling (PLS-SEM).* Sage publications, 2021.
53. J. Henseler, C. M. Ringle, and M. Sarstedt, "A new criterion for assessing discriminant validity in variance-based structural equation modeling," *J. Acad. Mark. Sci.*, vol. 43, no. 1, pp. 115–135, 2015, https://doi.org/10.1007/s11747-014-0403-8.

Text Classification: A Comprehensive Survey from Traditional Approaches to Deep Learning Methods

Ihab L. Hussein Alsammak, Wasan H. Itwee, Moamin A. Mahmoud, and Noor Islam Jasim

Abstract In the last decade, the number of texts and complex documents has increased exponentially, requiring a deeper knowledge of machine learning techniques for carefully classifying texts. The most fundamental task in natural language processing is Text classification. In natural language processing, many machine-learning algorithms have achieved excellence. The success of these approaches depends on their ability to understand nonlinear relationships and complicated models in data. In recent years, research in this area has seen an upsurge because of the unprecedented success rate of Deep Learning. Various datasets, evaluation metrics, and methods are suggested in the literature, so an up-to-date and comprehensive review is needed. In text classification tasks, Deep Learning-based models have outperformed traditional machine learning-based approaches in several areas (question answering, sentiment analysis, topic labeling, natural language inference, news classification, and named entity recognition). This paper discusses these tasks in detail, covering benchmark datasets and technical developments. This study eliminates the gap by providing an overview of the state of the art from 2014 to 2022, and the target is all tasks and models, from traditional to Deep Learning. Text classification tasks are thoroughly analyzed from over 70 articles, discussing technical contributions, strengths, and commonalities. This study compares various approaches, listing evaluation criteria along with their pros and cons. However, due to recent developments, emerging trends, and the rapid evolution of models, this study is influenced by specific datasets and metrics, leading to particular results. Future work

I. L. H. Alsammak (✉)
Ministry of Education, Directorate General of Education of Karbala, Karbala, Iraq
e-mail: ehablaith@gmail.com

W. H. Itwee
Karbala Sewerage Directorate, Karbala, Iraq

M. A. Mahmoud
Department of Computing, Institute of Informatics and Computing in Energy, College of Computing and Informatics, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

N. I. Jasim
College of Graduate Studies, Universiti Tenaga Nasional, Jalan Ikram-Uniten, 43000 Kajang, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_14

should include more recent research, broader datasets, further comparisons, and the exploration of new technologies in text classification.

Keywords Machine learning · Natural language processing · Text classification · Deep learning · Text classification tasks

1 Introduction

Natural language has been used by people to communicate since ancient times and is an effective tool to advance human society. People who communicate with each other are better able to express their views and understand those of others. The development of technology has made communication between humans and computers easier through various commands or codes. To understand natural language, computers must undergo special training. Spreadsheets and other structured data formats were recognized by older computers. However, in natural language, words are an unstructured data type whose meaning changes depending on the situation. Therefore, it is challenging for computers to understand the meaning of words. NLP (Natural Language Processing) is therefore required in such situations [1]. NLP is a branch of AI (artificial intelligence) and a field of computer science that deals with the interaction between human languages and computers. It also studies how to extract important information from natural language. In addition, NLP is a field of study that investigates and analyzes how natural language material fed into a computer can be modified and converted into a format more suitable for further processing. It is described as a spectrum of computer-based approaches for automated analysis and representation of human speech [2, 3]. Recent NLP strategies for dealing with unstructured textual input allow for high model performance with minimal data preparation, without sacrificing human interaction. For this reason, many ML (machine learning) methods have been proposed to classify text and include it in the services of most service providers. Text analysis usually employs a variety of methods to process the text, the most important of which is NLP. Principles of computational linguistics are applied to analyze lexical and linguistic patterns [4]. The development of NLP applications for English has made great strides. However, due to the rapidly growing availability of multilingual information on the Internet, the development of NLP applications for other languages has already begun [5]. TC the term “text” is used not only for basic categorization, but also for summary systems and question answering. The current response to textual information has long outpaced human solutions to these tasks, which means that approaches are not only interesting but also essential. Consequently, it is critical to developing accurate and unbiased text classification systems [6, 7]. Text classification tasks in NLP often used as benchmarks for evaluating new methods include Topic Labeling (TL), Sentiment Analysis (SA), Named Entity Recognition (NER), Natural Language Inference (NLI), News Classification (NC), and Question Answering (QA) [8–12].

1.1 Fundamental Contributions and Differences

Recently, several reviews have been published on TC. Four of them are reviews of TC and its subproblems [8, 9, 11–13]. In [12], the authors review TC algorithms using various text feature extraction techniques, basic model structures for text classification, dimensionality reduction strategies, and evaluation techniques. They then discuss the restrictions of every method and their application to real-world problems. In [11], the authors provide an overview of the latest DL-based TC methods, benchmark datasets, and evaluation metrics. The researchers shortened the overview to a brief definition of text tasks. About [8], this paper briefly discusses current TC methods, focusing on data flow from input text to output labels. Also pay attention to the differences between more current DL methods and older ones, both in terms of how they work and how they transform the input data. In [9], the authors conduct a comprehensive study and analysis of the primary models, covering both conventional and DL models, based on their structural frameworks. They also provide an introduction to the process and development of TC. They also list the essential details of DL models in expressions of basic model architecture. In contrast to previous reviews of text classification, we review current model inference—from traditional to DL—using studies from recent years. We focused on text classification tasks in detail, compiled various information in a table for each task with datasets, and reviewed relevant works. We also summarize various information tables and figures, including the required information for tasks with primary information datasets and a common benchmark for various applications, of state-of-the-art methods.

This study's main contributions are as follows:

- We give a thorough overview of more than 80 proposed models for TC.
- We have examined 57 common TC datasets.
- We offer an analysis of the effectiveness of a chosen group of 120 models across 6 tasks.
- We talk about a TC analysis with challenges.
- We present an overview of an analysis of methods in the literature for traditional methods with advantages, disadvantages, and applications.
- We also present the most effective methods from the literature for DL.

1.2 Organization of the Survey

The remaining sections of the study are categorized into the following: In Sect. 2 provides a summary of existing models related to TC tasks, including traditional and DL models with datasets, including summary tables for each task; Sect. 3 presents the existing classification techniques for traditional models (with advantages, disadvantages and applications) and DL models (the most effective methods). Section 4 summarizes the major challenges facing TC, including modeling challenges, data-driven challenges, and performance-based challenges. Finally, Sect. 5 provides the conclusion of the article.

2 Text Classification Tasks

This section presents six TC tasks discussed in this article, including 27 Topic Labeling, 44 Sentiment Analysis, 38 Question Answering, 20 News Classification, 17 Natural Language Inference, and 12 Named Entity Recognition, which are commonly presented as Text classification in many recent studies (see Fig. 1).

2.1 *Question Answering (QA)*

QA is a particular type of information retrieval. A QA system tries to determine the right response to the query posed in the natural language given a set of documents. Answering questions involves several disciplines. Information technology, language processing of natural language, artificial intelligence, knowledge and cognitive science and database management, are all involved. From a technical standpoint, knowledge representation and reasoning, information retrieval, natural or statistical language processing, and knowledge retrieval serve as possible building blocks for question answering [14]. They want to ask inquiries in their local language rather than being constrained by a particular query language, query formulation guidelines, or even a certain knowledge subject because they want responses that are brief and precise. The most recent method for matching user demands is to investigate the query from a linguistic perspective and make an effort to comprehend what the user means [15].

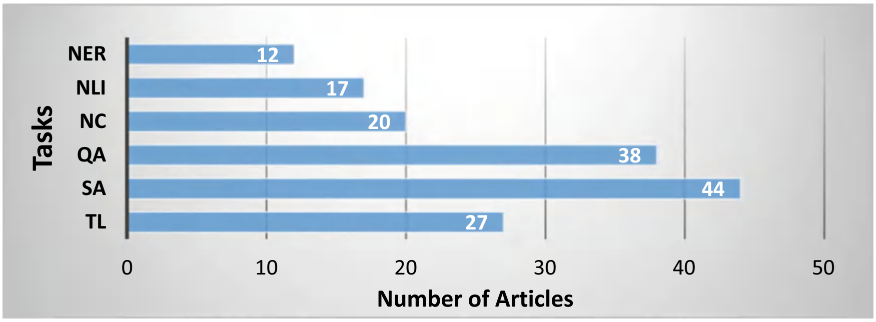


Fig. 1 TC tasks with the number of articles

2.1.1 Related Work and Statistics of Textual QA Datasets

Through previous studies shown in Table 1 and based on datasets as show in Fig. 2, we can give a summary of the contributions of researchers. For CNN and Daily Mail researchers such as [16], the stories are utilized as context to complete the fill-in-the-blank questions by using human-generated abstractive while one entity is hidden. The goal of NewsQA [17] is to provide a SQUAD-like QA dataset that is more challenging. The questions that the annotators are expected to come up with are based on CNN news, and the responses are spans of text that might be a person, place, sentence, number, etc.

Table 1 Statistics of QA datasets

Paper	Year	Dataset
[18]	2020	SST, MNLI, SQuAD
[19]	2019	RACE
[20]	2019	RACE
[21]	2018	Story Cloze Test
[22]	2018	Narrative QA
[23]	2017	TREC
[24]	2016	SST-1, Subj, TREC
[25]	2015	AG, Yelp P, DBPedia, AG News
[26]	2021	FinQA
[27]	2020	PIQA
[28]	2021	QASPER

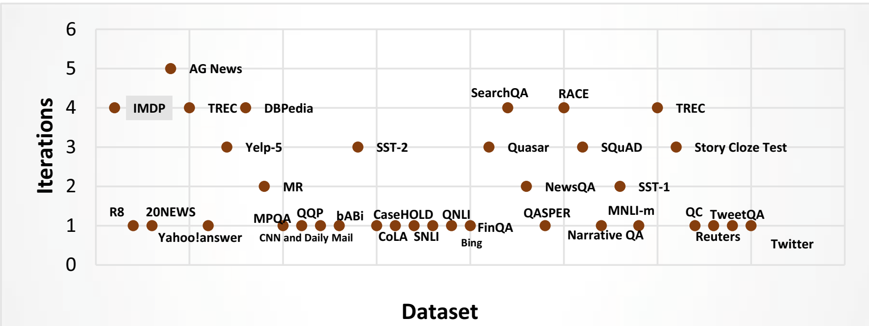


Fig. 2 The number of iterations for each database

2.1.2 Challenges and Chances

Question Answering Systems (QA) are complex systems comprising question processing, document processing, and response processing modules. The effectiveness of QA hinges on the proper functioning of these components, particularly the semantic categorization used to derive precise responses from document repositories. Textual QA, rooted in offline and online textual sources, remains the most extensively studied approach. It relies on datasets derived from either real test questions or annotated queries. While real test questions offer high-quality challenges, they are scarce and may not cover diverse QA requirements.

2.2 Sentiment Analysis (SA)

SA often known as Opinion Mining is a prominent NLP activity used to determine the orientation or sentiment of text supplied by users. Machine learning techniques are widely used to produce an accurate model, which is essential for performing sentiment analysis. It has been shown that deep learning and transfer learning approaches have higher usage and better outcomes, making them some of the most well-liked study topics worldwide. While experts utilize Twitter data and stock market news items to anticipate stock market trends, hotel and restaurant businesses examine reviews to gain a greater knowledge of the customs' requirements, preferences, and inclinations. Textual content is being generated in vast quantities across various platforms, including messaging services like Telegram and WhatsApp, social media sites like Facebook and Instagram, Google searches, news publishing sites, and many other sources. These platforms continuously produce massive volumes of text data every second.

2.2.1 Related Work and Related Work and Statistics of Textual SA Datasets

This study's goal is to evaluate several SA techniques that may be used as a guide in future research. Here, we have concentrated on datasets. As explained in Table 2 and as show in Fig. 3 through previous studies and based on datasets, we can give a summary of the contributions of researchers. An essential step in text data analysis is data preparation. Text data becomes increasingly complex as a result of the repetitions and redundancies in tweets, blogs, reviews, and other sorts of text. Data preprocessing is a filtering technique used in data normalization. The benchmarks that are utilized the most frequently in the literature are the datasets supplied by SemEval shared tasks. The workshop organizers compile and analyze user ratings from two industries, notably restaurants and computers. These datasets may be used directly for various tasks like aspect term extraction or aspect sentiment classification since they contain

Table 2 Statistics of SA datasets

Paper	Year	Dataset
[29]	2019	SemEval 2016 workshop
[30]	2019	Movie, food reviews (Amazon), Database (IMD)
[31]	2018	Social network sites
[32]	2022	TREC
[33]	2021	TREC, R8, R52
[18]	2020	SST, MNLI, SQuAD
[34]	2018	Dbpedia, SST-2, QQP, QNLI, CoLA
[23]	2017	TREC
[35]	2016	IMDB, Elec, RCV1, 20NG, Dbpedia, AG News
[24]	2016	SST-1, Subj, TREC
[36]	2015	SST-1, SST-2, QC, IMDB

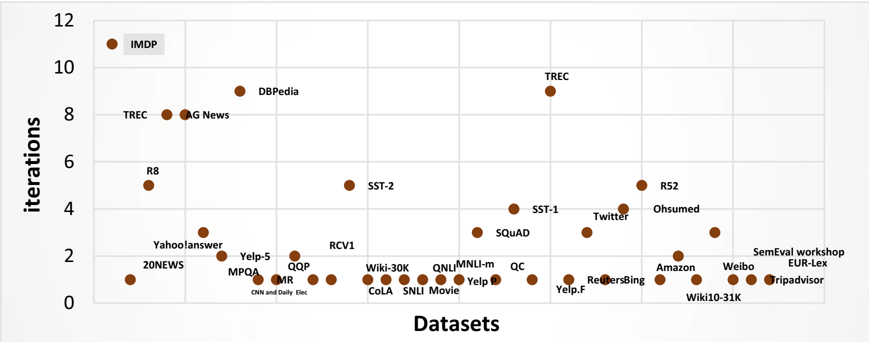


Fig. 3 Number of iterations for each database

annotations of aspect categories, aspect terms, and sentiment polarity (although not all of them do).

2.2.2 Challenges and Chances

Sentiment Analysis (SA) has evolved significantly since its inception in 2001, initially aimed at forecasting market sentiment from evaluative texts. The proliferation of Web 2.0 and user-generated content has expanded opportunities for research in SA across various domains. However, analyzing sentiment in social media posts remains challenging due to diverse textual features like punctuation, slang, emoticons, and URLs, compounded by nuances such as irony and ambiguity. Recent studies have focused on leveraging social media data for practical applications, including sentiment classification of reviews from platforms like TripAdvisor and Amazon, achieving high accuracies of 91.95% and 93.01% respectively.

2.3 News Classification (NC)

The development of news has significantly improved with the advancement of digital technology. The daily lives of people are significantly impacted by Internet news media platforms. Because internet news is frequently brief and incredibly informative, it can keep readers up to date. In the first half of 2018, there were 660 million Chinese network newsreaders, an increase of 160 million from the amount at the end of 2017. These readers made up 82.7% of all Internet users. One of the most common study topics is how to efficiently and reliably identify the massive number of news texts and extract important information. The web is enormous and continually updated. The number of news websites has expanded quickly in the digital age, and each one has a unique design and system for classifying content.

2.3.1 Related Work and Statistics of Textual NC Datasets

Through previous studies shown in Table 3 and based on datasets and as show in Fig. 4, we can give a summary of the contributions of researchers. The researchers [37] combined the many contextual representations that are already in use and discovered that knowledge graph-based document representations perform better. The researchers also point out that this is the first extensive analysis of the methodical integration of knowledge graph-based representations into the categorization of false news. Authors suggest a graph neural network-based approach that eliminates the requirement for feature engineering for fine-grained false news classification in order to collect this type of information for lengthy news pieces. Datasets were also accompanied by statistics, preprocessing instructions, and baseline models for monolingual and multilingual languages. The results of the tests demonstrate that embedding training in Kinyarwanda, a comparably more resource-rich language, results in effective cross-lingual transfer to Kirundi.

2.3.2 Challenges and Chances

Modern machine learning algorithms are developing into a workable answer to automatic processes like credibility detection by being able to analyze large sets of labeled and unlabeled textual inputs. However, one of the main issues are how to represent such data in a way that is appropriate for learning. In the past several years, significant progress has been achieved in this area, from massive, publicly available curated knowledge graphs to contextual language models that can distinguish between minute differences across a variety of texts. In literature, long news is classified to a significant extent, but work on news headlines is particularly scarce. Larger news articles are more difficult to classify since they need more statistical analysis and effort. Compared to those in news headlines, the chances of misclassification in the categorization of news stories with long descriptions are more apparent. Numerous scholars

Table 3 Statistics of NC datasets

Paper	Year	Dataset
[38]	2020	20NG, Ohsumed, MR
[39]	2020	R8, R52, Ohsumed
[40]	2020	20NG, R8, R52, Ohsumed
[41]	2019	DBpedia, 20NG
[42]	2019	20NEWS, OHSUMED, R8, R52
[43]	2017	IMDB, Dbpedia, RCV1
[44]	2022	PolitiFact, GossipCop
[45]	2021	AraCOVID19-MFH
[37]	2022	Social media
[46]	2021	N15News, AG News Fakeddit
[47]	2022	StockNet Astock

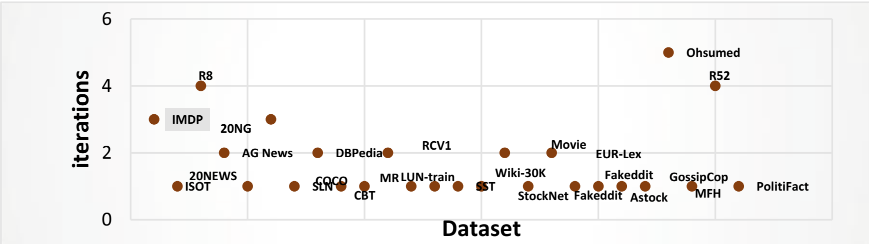


Fig. 4 Number of iterations for each database

have proposed methods for categorizing news headlines, which place each headline in its own predefined class. Finding the most likely words inside the class—words that are then deemed appropriate for that class—is how it is done. Then, all of those terms that are used often in all courses are eliminated.

2.4 Topic Labeling (TL)

TL is one of the classic examples of text classification. Authors define topic labeling (TL) as the task of recognizing one or more themes in a piece of text (i.e., its topics) and Other authors define topic labeling (TL) as one of several standard definitions for text classification (TC) tasks in the NLP research field, frequently used as benchmarks for evaluating new methods. Conversely, Topic Labeling (TL) refers to a topic analysis method that seeks to understand the text by identifying its complex themes. Topic labeling is a vital aspect of the topic analysis process, designed to assign one or more topics to each document, thereby enhancing the overall analysis. Is often known as “multi-labeled classification” [15, 48].

2.4.1 Related Work and Statistics of Textual TL Datasets

Through previous studies shown in Table 4 and based on datasets and as show in Fig. 5, we can give a summary of the contributions of researchers. For the Amazon dataset [49], to recall and score labels, Authors employed generative cooperative networks. The label recalling component of these networks creates positive and negative labels, while the label rating part separates positive labels from these labels. By feeding the same text representation through these networks during the label ranking portion of training, negative labels are sampled dynamically. Authors conducted extensive tests on several benchmark datasets to demonstrate HyperGAT's advantage over other cutting-edge approaches to the text classification problem [38]. Authors suggested using document-level hypergraphs to represent text documents, which increases the model's expressive ability and uses less computing resources. For the purpose of supporting representation learning on text hypergraphs, a principled model called HyperGAT based on a dual attention mechanism is presented. Authors of [39] introduced TextING1 for inductive text classification using GNN.

Table 4 Statistics of TL datasets

Paper	Year	Dataset
[49]	2021	AmazonCat-13K
[10]	2020	20NG, Reuters-2157
[50]	2020	SST2, IMDB, 20NG
[40]	2020	20NG, R8, R52, Ohsumed
[51]	2020	R8, R52, Ohsumed
[52]	2019	RCV1, EUR-Lex
[53]	2019	R8, R52, Ohsumed
[15]	2018	RCV1, NYTimes
[54]	2017	Twitter, AG, Bing
[24]	2016	SST-1, Subj, TREC
[25]	2015	AG, Yelp P, DBPedia, AG News

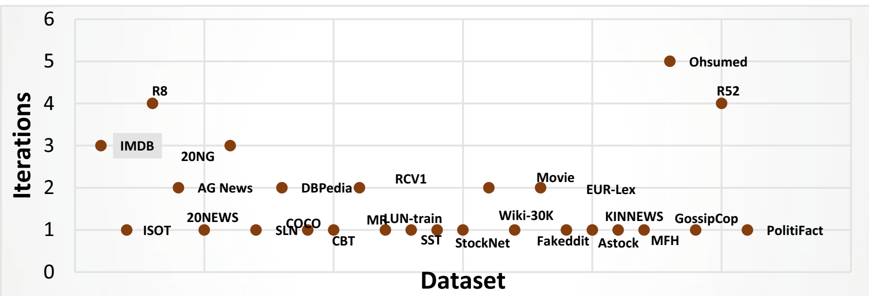


Fig. 5 Number of iterations for each database

2.4.2 Challenges and Chances

Although labeling text data takes a lot of effort, it is necessary for automatic text categorization. Especially when a very big quantity of data is required for training multi-label text classifiers, manually constructing numerous labels for each document may become problematic. A revolutionary multi-label active learning strategy can reduce the amount of labeled data needed without compromising classification accuracy, minimizing the need for human labeling. Only single-label issues can be handled by traditional active learning algorithms, meaning each piece of data can only have one label. The multi-label information is taken into consideration by new techniques, which then choose the unlabeled data that can most significantly reduce the projected model loss. For multi-label text classification, there are two approaches: Extreme Multi-label text Classification (XMC) and classical multi-label classification. The candidate label collection may be quite big, which increases computing complexity significantly in contrast to the classical multi-label classification issue. The goal of Extreme Multi-label Text Classification is to select the most pertinent labels from a very large label collection for each text. It is a very practical problem that has been widely implemented in several real-world contexts, such as proposing keywords to Amazon ads and categorizing Wikipedia articles with the most pertinent labels. The candidate label collection may be quite big, which increases computing complexity significantly in contrast to the standard multi-label classification issue [55, 56].

2.5 Natural Language Inference (NLI)

Humans conduct NLI based on extensive external information about language and the outside world. Comprehensive knowledge resources are essential in enhancing machines' ability to perform inference. Classifying all human knowledge and reasoning for broad inference remains an ongoing challenge. NLI is a classic text classification task, involving the determination of whether two sentences entail each other (classifying if the entailment occurs in one direction, both directions, or neither). In other words, It evaluates whether a 'hypothesis' is false (contradiction), true (entailment), or indeterminate (neutral) depend on a given 'premise'.

2.5.1 Related Work and Statistics of Textual NLI Datasets

The Natural Language Inference (NLI) job is frequently characterized as supervised pair classification of sentences. There has been a lot of recent work on developing explanations for classifier predictions on a single piece of text, but no attempts have been made to develop explanations for classifiers that operate on pairs of sentences. Natural Language Inference (NLI) large-scale datasets have made it possible to create several deep-learning models. Additional token-level data. Through previous studies

Table 5 Statistics of NLI datasets

Paper	Year	Dataset
[49]	2021	AmazonCat-13K
[57]	2020	IMDP, Yelp-5
[58]	2020	SST-5, MNLI
[59]	2019	TREC, Yahoo!answer
[60]	2019	Dbpedia, AG News, IMDP, Yelp-5
[34]	2018	Dbpedia, SST-2, QQP, QNLI, CoLA
[61]	2018	SQuAD, SNLI, SST-5
[62]	2019	ART
[63]	2016	Story Cloze
[64]	2010	Conversational Entailment
[65]	2019	SciTail, MNLI

shown in Table 5 and based on datasets, we can give a summary of the contributions of researchers. For NLI task might be beneficial for many downstream jobs, including the following applications: In retrieval-based question-and-answer (QA) systems, entailment functions are used as a search component to retrieve questions or answers that are implied by users' queries. The majority of NLI systems are built using machine learning techniques. Actually, the NLI issue is a supervised classification problem with two classes for binary decisions (infers, not infers) and three classes for three-way decisions (unknown or contradiction, infers, not infers). In the NLI problem, machine learning has been applied at several stages (preprocessing, comparison, and decision phases), typically in conjunction with other approaches. Machine learning methods concentrate on making predictions based on the characteristics discovered from training data (Fig. 6).

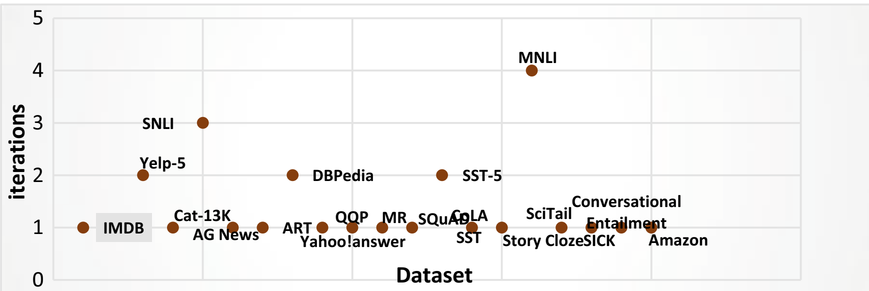


Fig. 6 Number of iterations for each database

2.5.2 Challenges and Chances

The Turing Test, proposed by Alan Turing to evaluate machine intelligence, has been criticized for its impracticality, encouragement of machine deception, and lack of incremental feedback mechanisms. As a more practical alternative, benchmark datasets have been developed, offering continuous numerical training, feedback and testing data, and evaluation frameworks for various linguistic tasks. Since the introduction of the RTE (Recognizing Textual Entailment) Challenges by the authors [66], numerous challenging NLI benchmarks have emerged. Originally, RTE tasks were significant reasoning challenges, driven due to challenges in semantic processing for applications such as question answering, machine translation, and information extraction. The number of benchmarks has surged recently. Earlier benchmarks, from around 2005–2015, were relatively small in size. In contrast, benchmarks developed after 2015 often contain significantly larger datasets, some exceeding 100,000 instances, which facilitates the application of DL techniques.

2.6 *Named Entity Recognition (NER)*

NER is NLP technique used for extracting information from unstructured text data sources such as emails, newspapers, blogs, and more. NER involves identifying nouns such as people, places, organizations, and other entities mentioned within a string of text, paragraph or sentence. An information extraction subtask called Named Entity Recognition (NER) groups proper names from unstructured texts into name categories. Many Natural Language Processing (NLP) subtasks, including text categorization, have evolved as a result of the NER task. Its goal is to try to find a way to improve the identification of named things that have been extracted. The goal of NER, a subtask of information extraction, is to identify and categorize text's atomic elements into specified groups, such as places, names of people, organizations, quantities, monetary values, percentages, and more. NER technology is employed in a wide range of applications, such as content classification for news providers, customer service, content recommendation engines, and search algorithms. These use cases are specific to particular domains, necessitating tailored word identification. For each specific requirement, a new NER model can be developed, or an existing model can be adapted to recognize particular terms [34].

2.6.1 Related Work and Statistics of Textual NER Datasets

Non-linear neural networks and word representations have been widely used in recent years to successfully handle a variety of NLP issues. Through previous studies shown in Table 6 and based on datasets and as show in figure, we can give a summary of the contributions of researchers. One of the earliest neural network models was presented by Authors in [61]. They showed that a feed-forward neural network combined with

Table 6 Statistics of NER datasets

Paper	Year	Dataset
[59]	2019	TREC, Yahoo!answer
[34]	2018	Dbpedia, SST-2, QQP, QNLI, CoLA
[67]	2014	Combination of Europarl, news commentary, IN
[61]	2018	Wikipedia, Common Crawl
[68]	2018	CONLL-2003

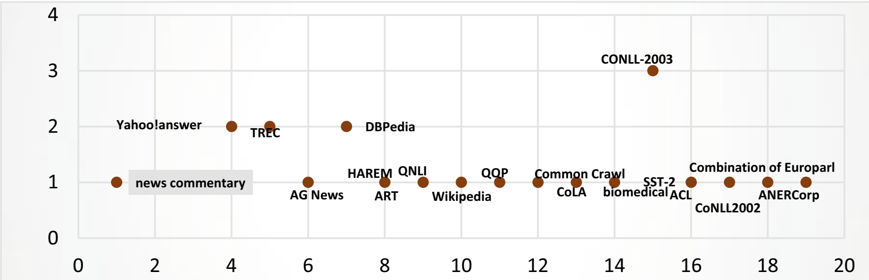


Fig. 7 Number of iterations for each database

a CRF would perform better than the majority of state-of-the-art methods in several NLP tasks, including NER, POS tagging, and semantic role categorization. Furthermore, recurrent neural networks (RNNs) have demonstrated their effectiveness in handling sequences of varying lengths by employing a recurrent hidden unit, the activation of which depends on each succeeding time step. Many RNN models, including LSTM (Long-Short Term Memory) and gated recurrent units, have shown considerable performance in modeling sequential data with practically no requirement for feature extraction or data reprocessing (Fig. 7).

2.6.2 Challenges and Chances

NER tools are evaluated based on criteria such as time efficiency, accuracy, and performance quality. Techniques like supervised, semi-supervised, and unsupervised learning are employed in NER, utilizing models such as SVM, CRF, and HMM. Libraries like SpaCy, Stanford NER, and TensorFlow offer NER models tailored for entities like people, organizations, and locations across different languages. Customizing NER models is essential for diverse applications, considering factors like prediction time, model size, and training simplicity.

3 Existing Classification Techniques

The cornerstone of the TC pipeline is selecting the optimal classifier. Without a comprehensive understanding of each classifier, accurately identifying the most effective model for a TC task is challenging. The following are some of the most commonly utilized techniques for text classification. Figure 8 illustrates the distinction between text classification using traditional methods and deep learning.

3.1 *Traditional Methods*

Conventional models increase the speed and accuracy of TC while expanding its scope. Preprocessing, which often includes statistics, word segmentation, and data cleaning, is the first step in preprocessing raw text data for training traditional models. In addition, text representation techniques such as Term Frequency-Inverse Document Frequency (TF-IDF) [69], Bag-Of-Words (BOW) [70], N-gram [71], word2vec [72], and Global Vectors for Word Representation (GloVe) [73] aim to represent preprocessed text in a way that is more accessible to computers while minimizing information loss. In this section, we present modern methods for text and document classification. In document classification, tree-based algorithms like decision trees (DT) and random forests (RF) are known for their efficiency and accuracy. Therefore, it can be concluded that combining two or more strategies can increase performance and provide better results. As explained in Fig. 9, the most commonly used models in Deep Learning include Recurrent Neural Networks (RNN), Feed Forward Neural Networks, Capsule Neural Networks, Convolution Neural Networks (CNN), and Attentional Mechanisms models. Figure 9 shows the number of items analyzed by model.

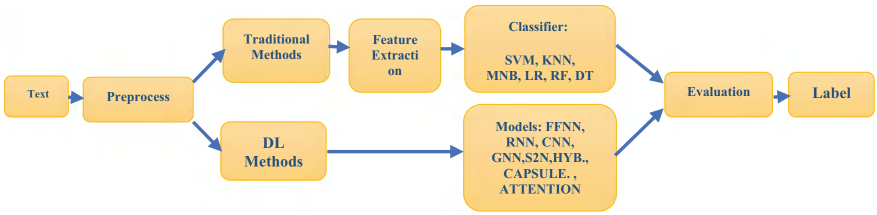


Fig. 8 Flowchart of the TC with classic methods and DL methods

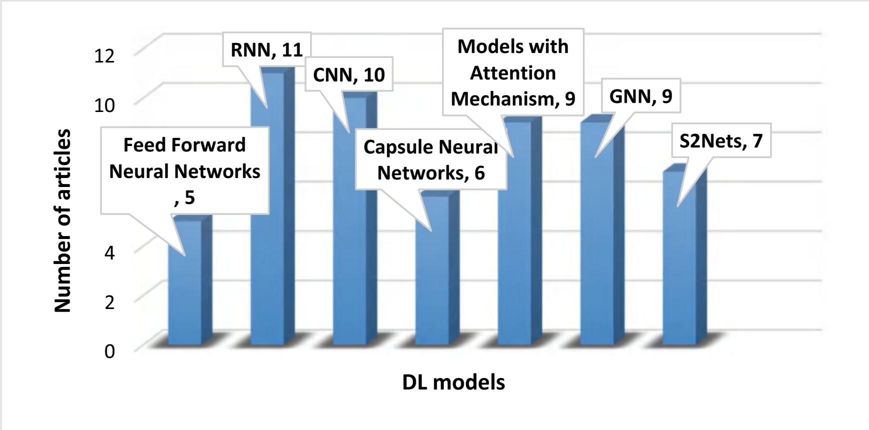


Fig. 9 Number of articles analyzed according to DL models

4 Text Classification Challenges

Text classification as a tool for effective information retrieval and evaluation is crucial for dealing with textual data. It is used for automatic classification and discovery of characteristic text types using NLP, ML, DM and other approaches. Text classification accepts many forms of text as input, and a vector. The vector is then sent to a Deep Neural Network (DNN) for training until the final condition is reached, and the downstream tasks check the performance of the training model. Although several recent TC models improve the accuracy index of most classification tasks, they do not show whether the machine "understands" the text at a semantic level as well as humans [9]. Furthermore, even a minor amount of noise in the sample can affect the decision confidence to fluctuate significantly or even reverse when the noise sample is present.

4.1 Modeling Challenges

In text classification, both classical and deep learning models like BERT excel. Enhancing computational power, data augmentation, and refining training processes are crucial. Challenges include efficient semantic-based text representation and model integration (RNNs, CNNs, attention mechanisms) for improved information extraction, highlighting ongoing opportunities for advancement.

4.2 Data-Driven Challenges

The following sections describe the technological challenges associated with the characteristics of these data (Zero-Shot/Few-Shot learning, Specialized, The task of classifying texts with multiple labels, The external knowledge).

4.3 Performance-Related Challenges

Although both the standard model and the DL model can perform well on most text classification tasks, their ability to avoid interference needs to be improved. It presents a technical challenge to achieve interpretability of the model for DL.

- The interpretability of the model: Deep neural networks excel in text classification, leveraging semantic mining and feature extraction.
- The semantic robustness of the model: Recent advancements in text classification have led to improved accuracy, yet models often struggle with adversarial samples.

5 Conclusions

Classification is a major challenge in machine learning, especially in the context of text and document datasets, which are continually expanding. The development and publication of supervised machine learning (ML) methods, particularly for text classification (TC), are crucial for enhancing document classification systems. This paper reviews over 70 text classification models developed in the past 8 years, significantly advancing the state of the art across various TC tasks. The review categorizes existing TC models from traditional to Deep Learning (DL). Traditional models enhance TC performance through optimized feature extraction strategies and classifier architectures. DL models improve performance by enhancing representation learning processes, refining model structures, and integrating new information and data. Additionally, the paper provides an overview of more than 57 established TC datasets and conducts a quantitative analysis of model performance against several publicly available benchmarks. Finally, it briefly discusses the ongoing challenges in TC. Advanced text classification significantly improves automated systems by enhancing response accuracy in question answering, aiding businesses in sentiment analysis, and organizing documents through topic labeling. It also boosts natural language inference for better understanding of sentence relationships, sorts and delivers relevant news, and extracts important entities with named entity recognition. These applications collectively advance data management, user interaction, and information processing across various domains. Future research should focus on integrating hybrid models that combine traditional feature extraction techniques with advanced deep learning methods to leverage the strengths of approaches, potentially improving text classification accuracy and robustness across diverse datasets. In Future work should include recent research, broader dataset analysis, deeper method comparisons, and exploration of emerging technologies in text classification.

References

1. Farkiya, A., et al., *Natural language processing using NLTK and wordNet.* IJCSIT Int J Comput Sci Inf Technol, 2015. **6**(6): p. 5465–5454.
2. Cambria, E. and B. White, *Jumping NLP curves: A review of natural language processing research.* IEEE Computational intelligence magazine, 2014. **9**(2): p. 48–57.
3. Palmer, M., *Natural language processing.* 2003.
4. Lavanya, P. and E. Sasikala. *Deep learning techniques on text classification using Natural language processing (NLP) in social healthcare network: A comprehensive survey.* in *2021 3rd International Conference on Signal Processing and Communication (ICPSC).* 2021. IEEE.
5. Gröndahl, T., *EAT: a simple and versatile semantic representation format for multi-purpose NLP.* arXiv preprint arXiv:1902.09381, 2019.
6. Howard, J. and S. Ruder, *Universal language model fine-tuning for text classification.* arXiv preprint arXiv:1801.06146, 2018.
7. Jalal, N., et al., *A novel improved random forest for text classification using feature ranking and optimal number of trees.* Journal of King Saud University-Computer and Information Sciences, 2022.
8. Gasparetto, A., et al., *A Survey on Text Classification Algorithms: From Text to Predictions.* Information, 2022. **13**(2): p. 83.
9. Li, Q., et al., *A Survey on Text Classification: From Traditional to Deep Learning.* ACM Transactions on Intelligent Systems and Technology (TIST), 2022. **13**(2): p. 1–41.
10. Wu, M., *Few-shot text classification with distributional signatures.* 2020, Massachusetts Institute of Technology.
11. Minaee, S., et al., *Deep learning-based text classification: a comprehensive review.* ACM Computing Surveys (CSUR), 2021. **54**(3): p. 1–40.
12. Kowsari, K., Jafari Meimandi, K., Heidarysafa, M., Mendu, S., Barnes, L., Brown, D.: Text Classification Algorithms: A Survey. Information, 2019. **10**(4).
13. Alsammak, I.L.H., et al., *A model for blockchain-based privacy-preserving for big data users on the internet of thing.* Indones. J. Electr. Eng. Comput. Sci, 2022. **26**(2): p. 974–988.
14. Yamada, I. and H. Shindo, *Neural attentive bag-of-entities model for text classification.* arXiv preprint arXiv:1909.01259, 2019.
15. Peng, H., et al. *Large-scale hierarchical text classification with recursively regularized deep graph-cnn.* in *Proceedings of the 2018 world wide web conference*. 2018.
16. Chen, M., et al., *Codah: An adversarially authored question-answer dataset for common sense.* arXiv preprint arXiv:1904.04365, 2019.
17. Rosemary Ke, N., et al., *Focused hierarchical RNNs for conditional sequence processing.* Proceedings of Machine Learning Research, 2018. **80**.
18. Lan, Z., et al., *Albert: A lite bert for self-supervised learning of language representations.* arXiv preprint arXiv:1909.11942, 2019.
19. Ran, Q., et al., *Option comparison network for multiple-choice reading comprehension.* arXiv preprint arXiv:1903.03033, 2019.
20. Zhang, S., et al., *Dual co-matching network for multi-choice reading comprehension.* arXiv preprint arXiv:1901.09381, 2019.
21. Sun, K., et al., *Improving machine reading comprehension with general reading strategies.* arXiv preprint arXiv:1810.13441, 2018.

22. Tay, Y., A.T. Luu, and S.C. Hui. *Multi-granular sequence encoding via dilated compositional units for reading comprehension.* in *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2018.
23. McCann, B., et al., *Learned in translation: Contextualized word vectors.* Advances in neural information processing systems, 2017. **30**.
24. Zhou, P., et al., *Text classification improved by integrating bidirectional LSTM with two-dimensional max pooling.* arXiv preprint arXiv:1611.06639, 2016.
25. Zhang, X., J. Zhao, and Y. LeCun, *Character-level convolutional networks for text classification.* Advances in neural information processing systems, 2015. **28**.
26. Chen, Z., et al., *Finqa: A dataset of numerical reasoning over financial data.* arXiv preprint arXiv:2109.00122, 2021.
27. Bisk, Y., et al. *Piqa: Reasoning about physical commonsense in natural language.* in *Proceedings of the AAAI conference on artificial intelligence*. 2020.
28. Dasigi, P., et al., *A dataset of information-seeking questions and answers anchored in research papers.* arXiv preprint arXiv:2105.03011, 2021.
29. Alharbi, A.S.M. and E. de Doncker, *Twitter sentiment analysis with a deep neural network: An enhanced approach using user behavioral information.* Cognitive Systems Research, 2019. **54**: p. 50–61.
30. Kraus, M. and S. Feuerriegel, *Sentiment analysis based on rhetorical structure theory: Learning deep neural networks from discourse trees.* Expert Systems with Applications, 2019. **118**: p. 65–79.
31. Tang, D. and M. Zhang, *Deep learning in sentiment analysis*, in *Deep learning in natural language processing*. 2018, Springer. p. 219–253.
32. Chen, Q., et al., *Dual Contrastive Learning: Text Classification via Label-Aware Data Augmentation.* arXiv preprint arXiv:2201.08702, 2022.
33. Yadav, R.K., et al., *Enhancing interpretable clauses semantically using pretrained word representation.* arXiv preprint arXiv:2104.06901, 2021.
34. Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* arXiv preprint arXiv:1810.04805, 2018.
35. Johnson, R. and T. Zhang. *Supervised and semi-supervised text categorization using LSTM for region embeddings.* in *International Conference on Machine Learning*. 2016. PMLR.
36. Liu, P., et al. *Multi-timescale long short-term memory neural network for modelling sentences and documents.* in *Proceedings of the 2015 conference on empirical methods in natural language processing*. 2015.
37. Koloski, B., et al., *Knowledge graph informed fake news classification via heterogeneous representation ensembles.* Neurocomputing, 2022.
38. Ding, K., et al., *Be more with less: Hypergraph attention networks for inductive text classification.* arXiv preprint arXiv:2011.00387, 2020.
39. Zhang, Y., et al., *Every document owns its structure: Inductive text classification via graph neural networks.* arXiv preprint arXiv:2004.13826, 2020.
40. Liu, X., et al. *Tensor graph convolutional networks for text classification.* in *Proceedings of the AAAI conference on artificial intelligence*. 2020.
41. Zhang, J., P. Lertvittayakumjorn, and Y. Guo, *Integrating semantic knowledge to tackle zero-shot text classification.* arXiv preprint arXiv:1903.12626, 2019.
42. Wu, F., et al. *Simplifying graph convolutional networks.* in *International conference on machine learning*. 2019. PMLR.
43. Miyato, T., A.M. Dai, and I. Goodfellow, *Adversarial training methods for semi-supervised text classification.* arXiv preprint arXiv:1605.07725, 2016.
44. Bhattarai, B., O.-C. Granmo, and L. Jiao, *Explainable tsetlin machine framework for fake news detection with credibility score assessment.* arXiv preprint arXiv:2105.09114, 2021.
45. Ameur, M.S.H. and H. Aliane, *AraCOVID19-MFH: Arabic COVID-19 Multi-label Fake News & Hate Speech Detection Dataset.* Procedia Computer Science, 2021. **189**: p. 232–241.
46. Wang, Z., X. Shan, and J. Yang, *N15News: A New Dataset for Multimodal News Classification.* arXiv preprint arXiv:2108.13327, 2021.

47. Zou, J., et al., *Astock: A New Dataset and Automated Stock Trading based on Stock-specific News Analyzing Model.* arXiv preprint arXiv:2206.06606, 2022.
48. Chen, J., et al. *Deep short text classification with knowledge powered attention.* in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2019.
49. Jiang, T., et al. *Lightxml: Transformer with dynamic negative sampling for high-performance extreme multi-label text classification.* in *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021.
50. Choi, S., et al. *Less is more: Attention supervision with counterfactuals for text classification.* in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP).* 2020.
51. Zhang, H. and J. Zhang. *Text graph transformer for document classification.* in *Conference on Empirical Methods in Natural Language Processing (EMNLP).* 2020.
52. Peng, H., et al., *Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification.* IEEE Transactions on Knowledge and Data Engineering, 2019. **33**(6): p. 2505–2519.
53. Huang, L., et al., *Text level graph neural network for text classification.* arXiv preprint arXiv:1910.02356, 2019.
54. Wang, J., et al. *Combining Knowledge with Deep Convolutional Neural Networks for Short Text Classification.* in *IJCAI*. 2017.
55. Johnson, R. and T. Zhang. *Deep pyramid convolutional neural networks for text categorization.* in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).* 2017.
56. Chen, K., et al., *Turning from TF-IDF to TF-IGM for term weighting in text classification.* Expert Systems with Applications, 2016. **66**: p. 245–260.
57. Zaheer, M., et al., *Big bird: Transformers for longer sequences.* Advances in Neural Information Processing Systems, 2020. **33**: p. 17283–17297.
58. Croce, D., G. Castellucci, and R. Basili. *GAN-BERT: Generative adversarial learning for robust text classification with a bunch of labeled examples.* in *Proceedings of the 58th annual meeting of the association for computational linguistics.* 2020.
59. Han, K., et al., *Delta: A deep learning based language technology platform.* arXiv preprint arXiv:1908.01853, 2019.
60. Yang, Z., et al., *Xlnet: Generalized autoregressive pretraining for language understanding.* Advances in neural information processing systems, 2019. **32**.
61. Sarzynska-Wawer, J., et al., *Detecting formal thought disorder by deep contextualized word representations.* Psychiatry Research, 2021. **304**: p. 114135.
62. Bhagavatula, C., et al., *Abductive commonsense reasoning.* arXiv preprint arXiv:1908.05739, 2019.
63. Mostafazadeh, N., et al., *A corpus and evaluation framework for deeper understanding of commonsense stories.* arXiv preprint arXiv:1604.01696, 2016.
64. Zhang, C. and J.Y. Chai. *Towards conversation entailment: An empirical investigation.* in *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing.* 2010.
65. Schmitt, M. and H. Schütze, *SherLIiC: A typed event-focused lexical inference benchmark for evaluating natural language inference.* arXiv preprint arXiv:1906.01393, 2019.
66. Marelli, M., et al. *A SICK cure for the evaluation of compositional distributional semantic models.* in *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14).* 2014.
67. Cho, K., et al., *On the properties of neural machine translation: Encoder-decoder approaches.* arXiv preprint arXiv:1409.1259, 2014.
68. Ghaddar, A. and P. Langlais, *Robust lexical features for improved neural network named-entity recognition.* arXiv preprint arXiv:1806.03489, 2018.
69. Baeza-Yates, R. and B. Ribeiro-Neto, *Modern information retrieval*. Vol. 463. 1999: ACM press New York.

70. Zhang, Y., R. Jin, and Z.-H. Zhou, *Understanding bag-of-words model: a statistical framework*. International journal of machine learning and cybernetics, 2010. **1**(1): p. 43–52.
71. Cavnar, W.B. and J.M. Trenkle. *N-gram-based text categorization.* in *Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval*. 1994. Citeseer.
72. Mikolov, T., et al., *Efficient estimation of word representations in vector space.* arXiv preprint arXiv:1301.3781, 2013.
73. Pennington, J., R. Socher, and C.D. Manning. *Glove: Global vectors for word representation.* in *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).* 2014.

Development of Smart Contract Blockchain Integrated Model (Lightblock) for IOT Applications

Mathuri Gurunathan, Moamin A. Mahmoud, Ihab L. Hussein Alsammak, and Saraswathy Shamini Gunasekaran

Abstract The Internet of Things (IoT) is rapidly expanding and evolving into a widely used computing service that requires substantial processing resources and storage for handling large amounts of data. Many IoT systems encounter inherent challenges such as resource constraints, limited storage capacity, and restricted connectivity, which often necessitate reliance on outsourced cloud services for storage and computation. Blockchain technology has emerged as a promising alternative due to its robust data security and privacy features. It facilitates automatic transaction processing and eliminates the need for intermediary verification, thereby significantly improving transaction speeds. Despite its advantages, blockchain faces challenges such as high computational demands and complexity, which can be incompatible with lightweight IoT devices that require scalable solutions. This paper examines how blockchain technology can be used to implement smart contracts in IoT applications. It focuses on creating a scalable and lightweight blockchain model called LightBlock. LightBlock is specifically designed to optimize the balance between security, privacy, and efficiency for IoT devices. The objective is to be reached by enhancing different components of an agile and flexible blockchain framework. The study is focused on enhancing security and privacy for users of IoT, guaranteeing complete security and usefulness. The evaluation of the discoveries was conducted using MetaMask, a web3 interface, and simulation tools. Subsequent research could include exploring different blockchain platforms and conducting real-world experiments with a wider range of IoT devices.

Keywords Energy efficiency · Lightweight blockchain · Internet of things · Security and privacy · Scalable blockchain

M. Gurunathan · M. A. Mahmoud · S. S. Gunasekaran
Department of Computing, Institute of Informatics and Computing in Energy, College of Computing and Informatics, Universiti Tenaga Nasional, 43000 Kajang, Malaysia

I. L. H. Alsammak (✉)
Ministry of Education, Directorate General of Education of Karbala, Karbala, Iraq
e-mail: ehablaith@gmail.com

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_15

1 Introduction

The Internet of Things (IoT) has been growing rapidly and evolving into a necessary computing service. This development needs a large amount of data storage and processing power [1]. However, many IoT applications are constrained by fundamental limitations such as resource constraints, storage capacity, and limited connectivity. As a consequence, it often rely on outsourced cloud services for storage and computation [2]. Although practical, this method brings certain security and privacy problems. As a result, Blockchain technology has emerged as a possible alternative because to its powerful data security and privacy features.

Blockchain, a peer-to-peer network-managed decentralized ledger, is revolutionizing transaction processing, data management, and a variety of services [3]. Blockchain presents a game-changing option for the energy industry and other Internet of Things applications, see Fig. 1. It enables automatic transaction processing, eliminating the need for intermediary verification and considerably increasing transaction speed. Additionally, the immutability of data recorded on the blockchain and shared by all network participants guarantees unparalleled data security and transparency as an outcome [4, 5].

Despite its potential benefits, Blockchain's complexity brings issues such as limited scalability, high bandwidth requirements, latency, and high energy consumption [6]. These elements make it computationally demanding and less suitable for Internet of Things applications. To address these issues, this study offers the Smart Contract Blockchain Integrated Components (LightBlock) approach. The integration of Blockchain technology into IoT applications is made easier by this architecture. This is achieved through the implementation of three crucial optimizations: Lightweight Consensus Algorithm (LCA), Scalable Blockchain Architecture (SBA), and a Throughput Management Scheme. Furthermore, the research entails the simulation of transactions using Solidity and the emulation of system performance using the NS3 simulator. By integrating these components and optimizations, the LightBlock model seeks to offer an effective approach for utilizing Blockchain technology in IoT applications. In the rapidly changing world of the Internet of Things (IoT), the energy industry or other IoT applications sector, making sure that data is accurate, secure, and automated has become a major challenge [7]. As IoT devices increase in number there is an increasing need to effectively manage and validate the data by these devices generate. Current centralized systems frequently fail to deal with a large amount of data, potentially leading to mistakes and vulnerabilities. Moreover, the energy industry encounters various challenges in areas like billing, grid management, and verifying transactions. Conventional approaches are often manual, labor-intensive, and error-prone, resulting in inefficiencies, disputes, and extra costs. To overcome these issues, integrating IoT technology with blockchain-based smart contracts is crucial. IoT devices can collect vast amounts of data, including energy consumption patterns, device conditions, and environmental factors [8]. However, a key challenge remains in ensuring the data's integrity, authenticity, and confidentiality.

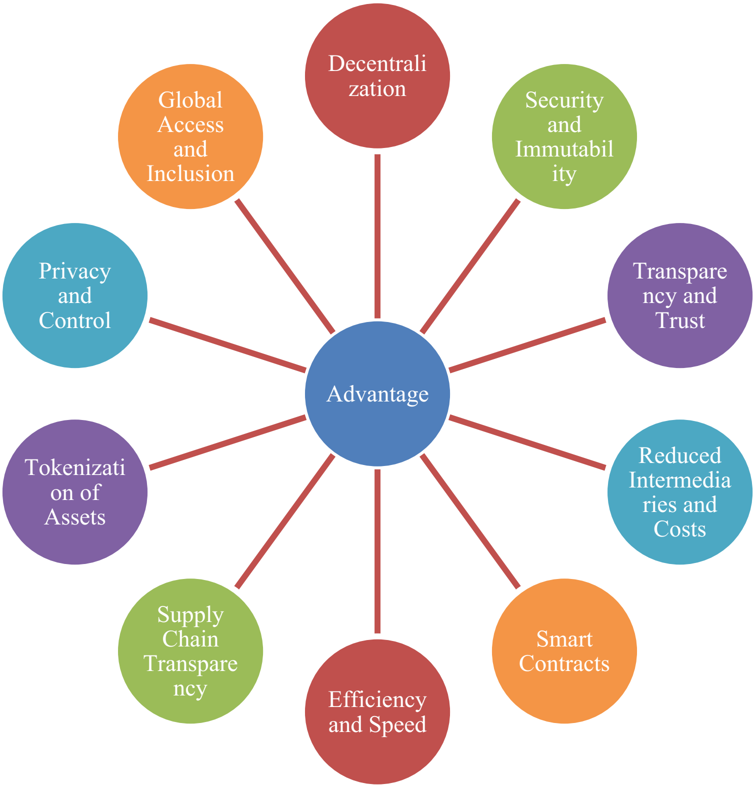


Fig. 1 Advantage of blockchain technology

Blockchain technology offers a decentralized and unchangeable ledger, creating a secure platform for recording and verifying data generated by IoT devices [9]. However, the inherent complexity of blockchain introduces challenges like limited scalability, significant bandwidth demands, latency issues, and high energy consumption. Integrating smart contracts, which are self-executing agreements based on preset rules, can help address these challenges [10]. This integration allows for the automation of tasks such as billing, grid management, and device authentication, reducing manual effort while enhancing overall accuracy and operational efficiency.

However, implementing an integrated IoT and smart contract-based solution for the energy sector or other IoT applications requires careful consideration of multiple factors, such as scalability, interoperability with legacy systems, regulatory compliance, and the creation of user-friendly interfaces for all stakeholders [11]. In addition, addressing energy efficiency and cost-effectiveness in the deployment of IoT devices remains an immense challenge for researchers.

Consequently, the primary objective of this research initiative is to develop a flexible and effective IoT and smart contract-based solution for the IoT applications sector. This solution seeks to revolutionize data management, verification, and automation processes, establishing new standards for operations' reliability and efficacy. The research aims to demonstrate the viability and benefits of this integrated approach in real-world applications through testing and validation. The scope of this study includes developing a specialized Scalable Lightweight Blockchain Integrated Components (LightBlock) model that addresses the specific requirements of IoT devices, with a primary focus on assuring end-to-end security. The study includes the development of a comprehensive and adaptable model for a lightweight, scalable blockchain, carefully balancing the requirements of IoT applications while maintaining the highest levels of security, privacy, and operational efficiency. In addition, include thorough performance testing of LightBlock, using Smart Contract Solidity on the Ethereum platform in conjunction with Discrete-event Network Simulator (NS3) for simulation scenarios.

2 Literature Review

The goal of the Internet of Things (IoT) is to connect control components and billions of electronic sensors to create interconnected systems that improve efficiency, convenience, and global living standards across all IoT areas [12]. Despite the potential benefits, confidence in IoT devices has been shaken by various incidents. Security worries have become a major issue due to the exponential growth of data produced by numerous sensors. These sensors often have power management capabilities, memory and limited storage, highlighting the critical importance of information security and protection [13, 14]. In response to these challenges, analysts recommend creating lightweight security solutions specifically designed for IoT devices. The term "lightweight" refers to efficient algorithms that require minimal resources, prioritizing robustness over compromised security implementations [15–17].

Blockchain (BC) is increasingly being integrated with IoT [18, 19], even though other security methods like physical layer security are also available [20–22]. BC functions as a distributed ledger for transactions, facilitating data exchange and storage, including personal information, contracts (like Ethereum), and currency (like Bitcoin). Its strong security and privacy features have garnered considerable attention from both scholars and industry professionals across various fields. The decentralized nature of BC removes the need for a central authority when adding new blocks, as all network nodes must validate and approve each addition. This feature not only boosts security but also ensures transaction transparency. Adding new blocks involves solving intricate puzzles through mining and consensus algorithms, engaging the entire network in the process. This makes BC a flexible solution for enhancing security and confidentiality across various domains. However, IoT devices typically rely on traditional cryptography due to significant resource limitations, such as restricted processing power, bandwidth, and storage. These constraints present challenges for effectively implementing complex security measures within IoT environments.

Additionally, a significant number of IoT devices depend on centralized communication models, where cloud servers handle the identification, authentication, and connectivity of all devices. This centralization creates a single point of failure: any issues at the central hub can jeopardize the entire network, potentially resulting in widespread system outages. Moreover, privacy protection for devices and their communications typically relies on traditional methods, which may not be sufficient to meet the growing and evolving privacy challenges in today's advanced IoT systems.

However, implementing blockchain (BC) in IoT comes with several challenges. One of the primary issues is the complexity of consensus algorithms. Algorithms such as Proof of Work (PoW) and Proof of Stake (PoS) demand substantial computational power, which is often beyond the capabilities of IoT devices with limited processing resources. Alongside its benefits, blockchain also encounters significant hurdles in terms of scalability and overhead. In blockchain networks, each new block must be distributed to each node for confirmation which increases scalability requirements and adds considerable overhead, ultimately reducing efficiency in IoT applications. Latency is another major concern; the time it takes for nodes to confirm and add new blocks can be lengthy. This delay is particularly problematic in IoT systems, where real-time data transmission is essential. Additionally, throughput is a major concern. Blockchain networks have restricted transaction processing capabilities; for example, Bitcoin can manage only around seven transactions per second. This limitation creates difficulties for IoT applications that typically need fast and frequent transactions. The constraints related to scalability, latency, and throughput underscore the persistent challenges of effectively integrating blockchain technology into IoT environments.

3 Methodology

3.1 Smart Contract System Structure

Smart contracts are used in various industries, including energy, supply chain management, real estate, and finance. For instance, in cryptocurrency transactions, a smart contract can automatically transfer funds to a seller once the buyer confirms receipt of the goods. This method offers a transparent, efficient, and trustless way to execute transactions, transforming how agreements are established and executed.

An example depicted in Fig. 2 illustrates a smart contract facilitating a transaction between a buyer and a seller, finalizing the transaction upon payment confirmation. In an IoT topology, Party 1 communicates with an IoT device to gather information or trigger actions. The smart contract, deployed on the blockchain, governs agreements and transactions as an automated middleman, ensuring transparent and

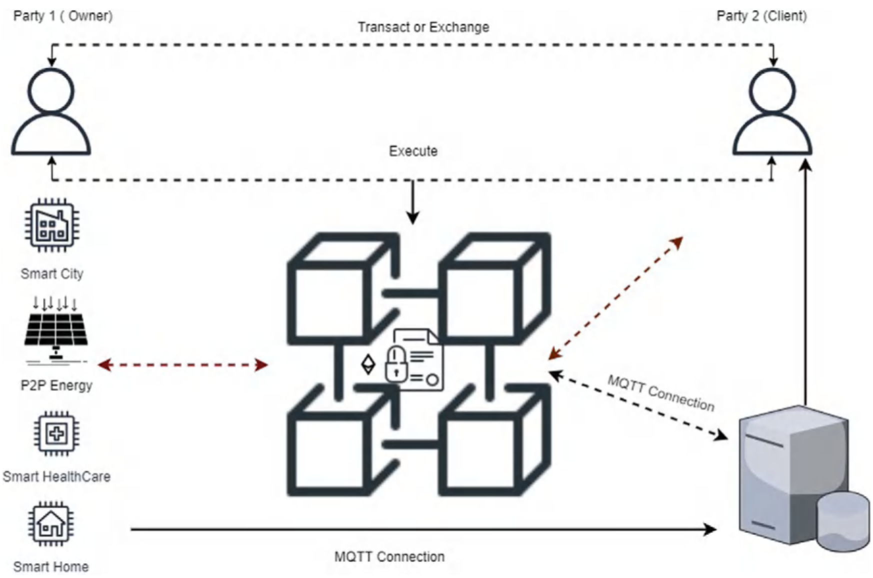


Fig. 2 IoT topology overview of using smart contract

trustworthy interactions by enforcing predefined conditions. This process integrates MQTT messaging with smart contracts, enabling secure, automated, and transparent IoT interactions.

Smart contracts on blockchain execute predefined conditions automatically using "if/then" logic, ensuring secure, tamper-proof transactions. Parties agree on terms, which are coded into the contract. Businesses offer tools to simplify this process. The blockchain updates post-transaction, allowing only authorized users to view results. Figure 3 show the smart contract execution process of the diagram. Developing decentralized applications (Dapps) involves using blockchain technologies like Solidity, Truffle suite, Remix IDE, Web3.js, and Metamask. These tools facilitate creating, testing, and interacting with smart contracts on the Ethereum network, ensuring transparency, security, and decentralized control. Figure 4 illustrates system communication with smart contracts and IoT devices.

Solidity is a high-level, statically-typed programming language designed for creating smart contracts on blockchain platforms, particularly targeting the Ethereum Virtual Machine (EVM). The language facilitates defining contracts with state variables, functions, and event declarations.

1. Creation of a Contract: a simple contract with a state variable and a function to set its value.
2. Event Declaration: an event declaration to transmit information to external consumers.
3. Constructor: a constructor method initializing the contract's state variable upon deployment.

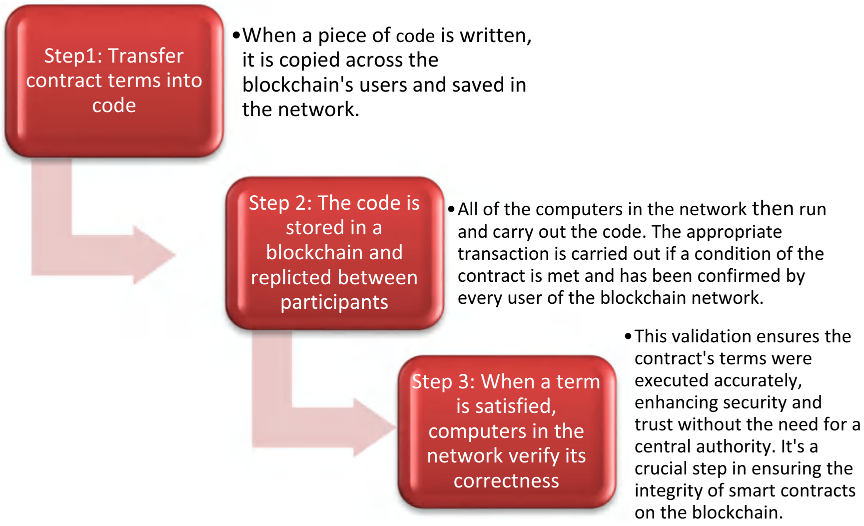


Fig. 3 Smart contract execution process

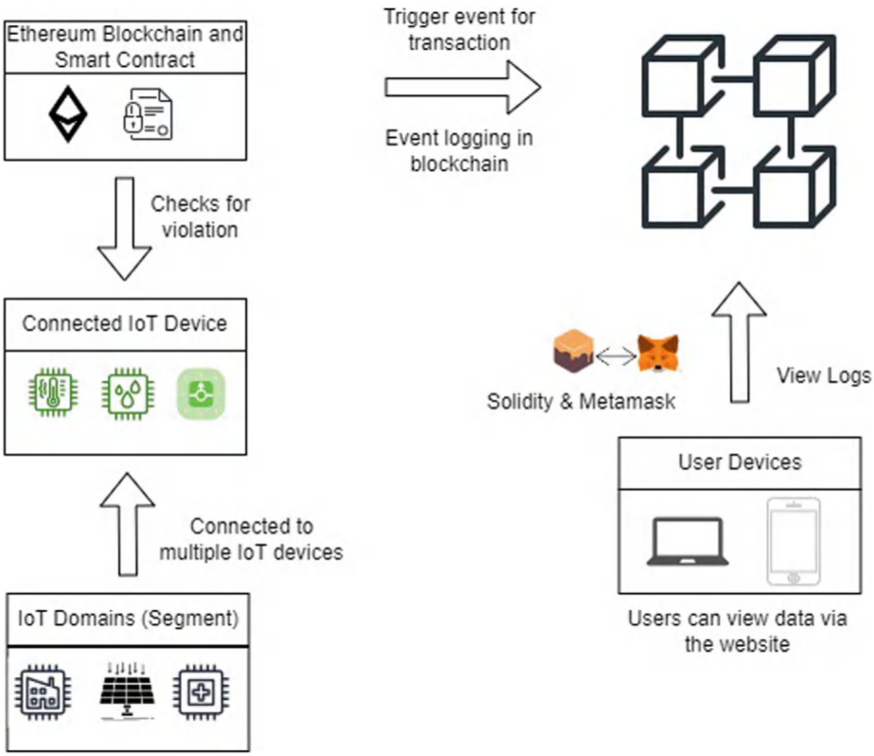


Fig. 4 System communication with the smart contract blockchain

4. Mint and Send Function: a token contract with minting and sending functionalities.

The Truffle Suite is a comprehensive framework for Ethereum blockchain development, including tools like Truffle for smart contract lifecycle management and Ganache for local testing. Remix IDE, a web-based environment, supports Solidity for creating, testing, and deploying smart contracts with integrated debugging tools. Web3.js, a JavaScript library, connects web applications to Ethereum, enabling transaction execution and smart contract deployment. MetaMask, a browser extension, functions as a digital wallet for Ethereum-based assets, enhancing accessibility and security. NS3, a network simulation tool, evaluates blockchain systems in device management, simulating interactions between IoT devices and the blockchain to ensure performance and scalability.

3.2 System Structure for Smart Contracts

The Ethereum blockchain was chosen for implementing the proposed blockchain-based smart contract due to its robust development tools, comprehensive documentation, and active community. It supports highly complex smart contract technologies, offering flexibility in implementation processes. Ethereum enables the creation of user-friendly applications atop blockchain, facilitating interaction with smart contracts and data. Smart contracts in Solidity are executed on the Ethereum Virtual Machine (EVM), tailored for decentralized, account-based transactions. Key tools include Remix IDE for seamless smart contract development, Ganache-cli for local blockchain emulation, and Web3.js for Ethereum interaction. These tools support Solidity development, deployment, and testing, ensuring accurate contract execution and evaluating system performance. Smart contracts facilitate secure data trading among customers (or miners) through a transparent, conflict-free system, preventing unauthorized access. Initially, customers access a trusted ledger on the blockchain for data exchange among nodes. These contracts, developed with intricate JavaScript logic, ensure sensitive data security in the blockchain database, accessible only to authorized users. The construction flow and interaction of the smart contract with SCADA systems. Using Remix IDE in the Solidity environment, a data monitoring mechanism is created. The system integrates the Ethereum blockchain, a local web server hosting a front-end application, and smart contracts deployed on the blockchain, requiring nodes to connect via the same chain ID for unified network operations.

4 Conceptual Model/Features of Lightblock

The goal of the lightweight blockchain study is to develop and standardize robust protocols and practices for their operational environments. No single model meets all criteria, as noted in [23]. Therefore, we propose a method that combines established blockchain threat defenses with the optimization and integration of four components—SBA, LCA, GCCF, and TMS. To evaluate the model's suitability for IoT applications, we conduct a case study involving a smart home, assessing resource usage and throughput. This case study involved connecting smart devices through the Internet in order to achieve optimal efficiency and resource utilization. A simulation study was conducted using a discrete-event network simulator to assess both security (through testing for denial of service, dropping attacks, and blockchain attacks) and efficiency (by analyzing metrics such as latency, bandwidth overhead, processing time, and scalability). Figure 5 illustrates a conceptual model highlighting the study's four key aspects.

The conceptual model of LightBlock is made up of three key elements that work together to improve its speed and efficiency. First, it strategically reduces network

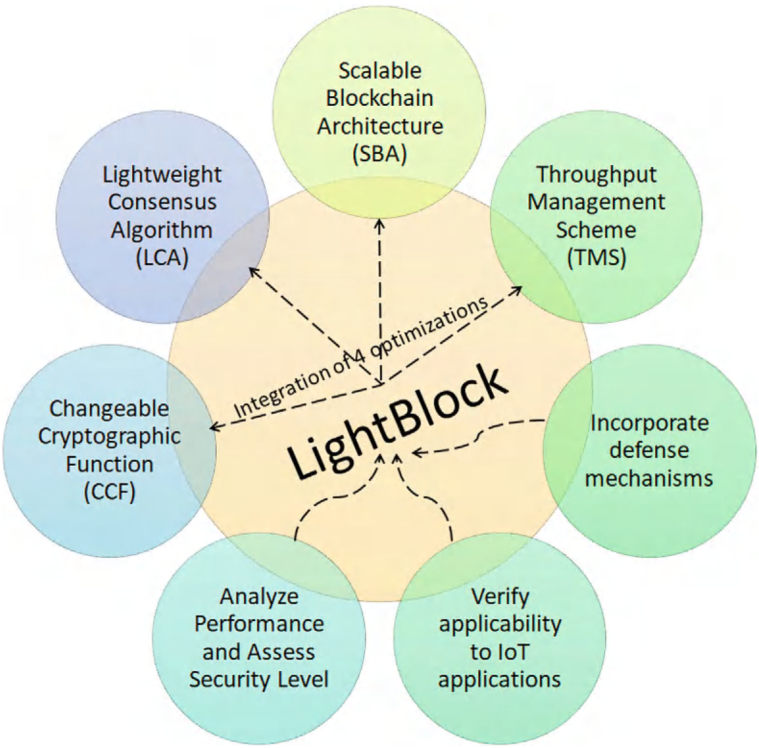


Fig. 5 Conceptual of proposed model

overhead by implementing Ethereum's Proof of Stake (POS) consensus process. By eliminating the traditional Proof of Work (POW) method, this solution reduces the amount of computing power needed for transaction validation, which could improve scalability and save energy. Secondly, LightBlock's dedication to greater energy efficiency is reinforced by the inclusion of Ethereum 2.0 (ETH 2). A major improvement, ETH 2 moves from Proof of Work (POW) to a more environmentally friendly consensus process. This change has the potential to significantly increase the network's overall sustainability.

The integration of Decentralised Applications (DApps) with the Ethereum network, last but not least, emphasises the need for increased throughput. DApps are a good option for apps with high transaction volumes as it can handle several transactions at once. By utilising Ethereum's well-established infrastructure, LightBlock may have access to a widely-used platform that will ease the burden on the network. Overall, this conceptual model fits in well with current trends in blockchain technology because it puts scalability, sustainability, and operating efficiency at the top of its list of priorities. The LightBlock conceptual model is carefully crafted to integrate Scalable Blockchain Architecture (SBA), a Throughput Management Scheme (TMS) and a Lightweight Consensus Algorithm (LCA). These components work together to improve the system's performance and efficiency, as detailed in the previous section.

i. (SBA): This implementation explores various network topologies such as Peer to Peer (P2P), Star, Bus, Ring, and Mesh to identify effective structures for high scalability, low overhead, and fast processing. P2P architecture enables direct communication between nodes, enhancing decentralized interactions. While reducing direct interactions can cause some latency and overhead, it significantly improves scalability and supports the decentralized nature of blockchain.
ii. (LCA): Several consensus algorithms are considered, including PoW, PoS, and DPoS. The Proof of Stake (PoS) mechanism is crucial for building a lightweight blockchain for smart contract systems. PoS is more environmentally friendly, reducing resource consumption and energy use by relying on validators' stakes rather than processing power. This results in lower transaction fees and faster confirmations, enhancing user experience. PoS ensures security through financial incentives, as validators risk financial losses for dishonest behavior. While PoS is pivotal, additional optimizations like smart contract efficiency and Layer 2 solutions are essential for full scalability and efficiency.
iii. (TMS): When the transaction volume exceeds the maximum, transactions are added to a queue. Some transactions, however, require immediate processing. Prioritizing these transactions helps manage throughput effectively. Utilizing asynchronous processing with events and callbacks allows concurrent execution, improving overall throughput.

5 Proposed Model Development

This section provides an Overview of system components and development tools for blockchain implementation and smart contracts. The Ethereum blockchain structure was selected for implementing the proposed smart contract system due to its robust development tools, comprehensive documentation, and active community support. Ethereum supports highly complex smart contract functionalities, allowing for flexible implementation processes.

Applications can interact with Ethereum through user-friendly interfaces, enabling widespread adoption of smart contracts. Solidity, the language for Ethereum smart contracts, facilitates decentralized and account-based transactions. Key development tools include Remix IDE for seamless contract development and debugging, Ganache-cli for local blockchain simulation, and Web3.js for Ethereum API interaction. These tools, integrated with an IDE, enable Solidity contract development, deployment, and testing, ensuring reliable contract execution and assessing system performance metrics (Fig. 6).

Smart contracts facilitate secure and transparent data trading among customers (or miners) on a blockchain-based system, preventing unauthorized access. Initially, users access a single trusted ledger sourced from the blockchain, with data exchange occurring directly between nodes. These contracts, developed with intricate JavaScript logic, ensure sensitive data remains securely stored in the blockchain database, accessible only to authorized parties. The blockchain-based smart contract system, depicted in Fig. 7, comprises the Ethereum blockchain, a front-end application hosted on a local web server, and smart contracts deployed on the blockchain. Nodes join by installing an Ethereum client and connecting to a unified chain ID or network ID, enabling transaction execution, mining, and participation in the consensus process. Smart contracts serve as decentralized business logic for Dapps, distinguishing them from centralized applications. Contract deployment involves using an Ethereum account to publish contracts, which become universally accessible across network nodes upon validation of the deployment transaction by miners. For user interaction, browsers or the MetaMask plugin for Chrome alongside an

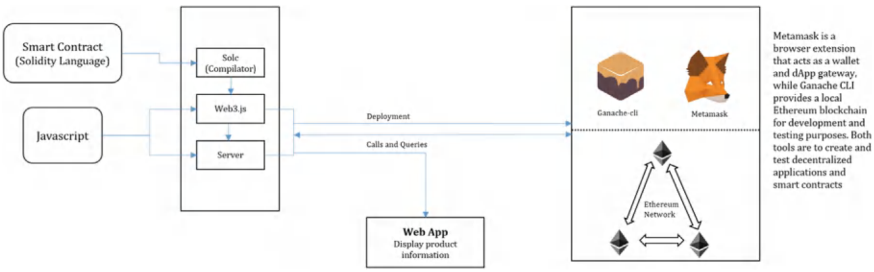


Fig. 6 Illustrates the framework used for creating, compiling, deploying, and testing smart contracts, employing Truffle and Ganache-cli

Ethereum account are essential. MetaMask acts as a secure cryptocurrency wallet, enabling connection to blockchain-based applications on a decentralized web platform. It allows users to import their Ethereum accounts, effectively transforming their systems into virtual blockchain nodes. Additionally, MetaMask injects web3.js into loaded HTML pages, serving as a wrapper library for encoding, decoding, and facilitating interaction with the blockchain's RPC port. The Application Binary Interface (ABI) standardizes communication with contracts in the Ethereum environment. Smart contract creation involves three stages: Solidity programming for coding, compiling with solc, and contract deployment on the network. The Data Collection Layer within the blockchain system integrates all nodes, enabling connections among IoT devices and facilitating event-triggering mechanisms to activate node actuators or retrieve sensor data. Contract accounts within this layer execute bytecode on the Ethereum Virtual Machine (EVM), manage contract addresses, store ether holdings, and maintain permanent data on the blockchain.

Users interact with the Ethereum network by selecting and connecting to their preferred network, unlocking accounts for transactions, and integrating web3.js into web pages for blockchain interaction via JSON-RPC. Graphical user interface (GUI) components employ Bootstrap for web interface construction, utilize web3.js for seamless blockchain interaction, and leverage truffle-contract for advanced transaction functionalities and synchronous control until mining completion. The Contract

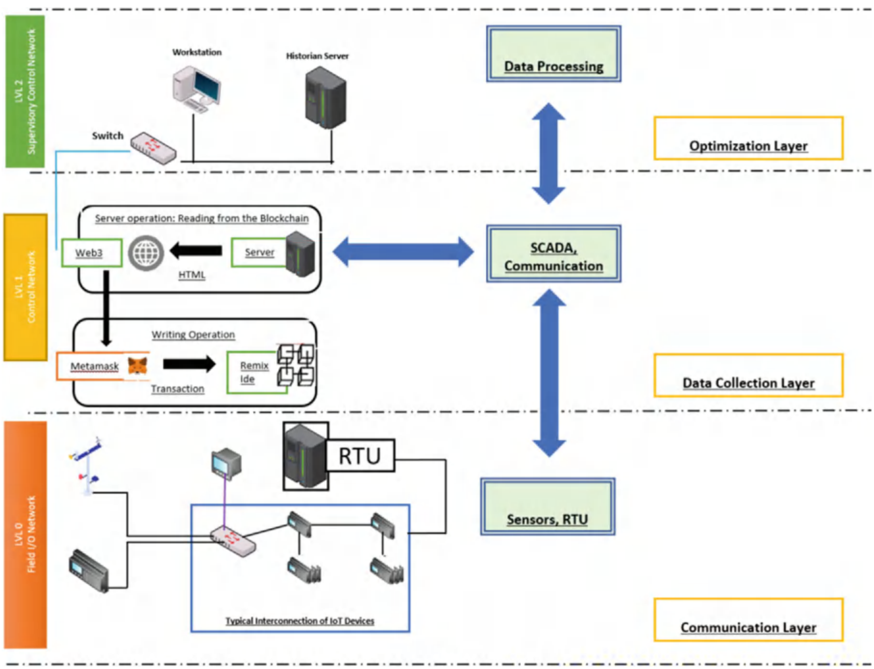


Fig. 7 System structure for smart contracts

ABI, a JSON-formatted specification detailing contract functions and arguments, facilitates integration into front-end applications without necessitating blockchain storage. Typically obtained post-compilation in formats like contractName.json using the Truffle framework, it ensures standardized communication with deployed smart contracts. Ethereum's toolset is tailored for constructing decentralized applications (Dapps) that interact with specific smart contracts on the blockchain using ETH for transactions or function calls. These tools are adaptable for applications connecting to private Ethereum blockchains, as envisioned in our smart contract system. A web application was developed to access and display data stored in our blockchain smart contract system, aligned with the objectives outlined in this dissertation. Contract deployment on a test blockchain utilized tools detailed in Fig. 7, including: web3.js: Ethereum JavaScript API facilitating smart contract testing and web development, enabling contract access and function invocation. MetaMask: Browser add-on ensuring secure Ethereum account management, facilitating seamless connection to diverse blockchain networks and account management. Http-server: Simplified tool supporting local HTTP server creation for efficient application development purposes.

The project structure comprises two main directories: "client" and "smart contract". The "client" directory hosts a React application, while "smart contract" contains files for the smart contract code and deployment script. Setting up the React application involves using npm to install dependencies like the "express" package, which is added to the project's package .json file. Writing Solidity involves crafting a smart contract named "Transactions". This contract allows deposits, withdrawals, and balance inquiries. It includes a SPDX license identifier indicating it's released under the GNU General Public License 3.0. The deposit function is declared as external and payable, enabling it to receive Ether from external calls. Similarly, the withdraw function sends Ether to a specified address using the transfer method, while getBalance returns the contract's current Ether balance and getAddress provides the contract's address.

Deploying the smart contract utilizes tools like Ganache, MetaMask, Remix, and a web server. Upon execution, the script interacts with the Ethereum network to deploy the contract, generating a transaction logged in the terminal along with the deployed contract's address. This deployment enables decentralized functionality, allowing the smart contract to interact with other contracts and network participants. To execute the deployment script, MetaMask facilitates creating an Ethereum account for transactions on the Ganache test network, requiring ETH to pay deployment gas fees. This process establishes connectivity to the Ethereum network, crucial for deploying and managing smart contracts through nodes accessing the blockchain for communication.

Deploying a smart contract and integrating it with a web application involves several key steps that enable interaction between the application and the Ethereum blockchain. This process typically begins with setting up necessary files and configurations within the project structure. Upon deployment of the smart contract, additional components like a new utils folder and constants .js file are created in the src folder. The constants .js file houses critical details such as the contract's address and its Application Binary Interface (ABI), which is automatically generated during deployment. The ABI serves as a crucial intermediary that defines how the application can communicate with the deployed smart contract. To initiate interaction with the smart contract from the web application, the first requirement is connecting to MetaMask, a popular Ethereum wallet browser extension. This connection enables the application to gain access to the user's Ethereum account, facilitating transaction signing and submission to the blockchain. The architecture typically involves:

- Provider and Signer Setup: The provider acts as the gateway to the Ethereum blockchain, granting the application access, while the signer represents the Ethereum account authorized to sign transactions.
- Contract Instance Creation: Using the imported ABI and contract address from constants .js, the application creates an instance of the smart contract. This instance is crucial for executing functions defined within the smart contract on the blockchain.
- Functionality Segmentation: The application's functionality can be segmented into read and write operations:
- Read Functions: These functions query data from the smart contract without altering its state. Examples include retrieving account balances or other stored data.
- Write Functions: These operations modify the contract's state, typically involving actions like depositing or withdrawing Ether. Each function requires proper authorization and security checks to ensure secure interaction.

Implementation Details: Key functions include:

- Connecting to MetaMask: Facilitates access to the user's wallet.
- Connecting to Smart Contract: Establishes communication with the deployed smart contract using its ABI and address.
- Reading.

6 Results and Discussion

The proposed blockchain architecture was tested by executing smart contract functions redundantly for parallel processing. This section evaluates CPU and memory performance with different consensus algorithms (PoW and PoS) to determine Ethereum's suitability for IoT applications and identify the best consensus mechanism for resource-limited environments. The architecture meets strict security standards, ensuring data availability, integrity, and tamper-proof confidentiality. For added security in data management and transmission, users can securely store their private keys using network-based methods.

The first phase of testing a new device involves examination, block extraction, validation, block revision, and finalization, which are essential steps in the device registration process. Following validation, the storage node manages block revision to ensure there is adequate space for adding confirmed blocks, which are then incorporated into the record.

When a user submits a request, a verification block is generated and checked by a distributed control authenticator. Upon successful verification, the user gains access to the blockchain network. Hash computations are performed on the blockchain ledger, with new blocks requiring resources and verification through hash-based functions. Resources are allocated for approved blocks. Figure 8 shows the testing process with and without fault tolerance using simulation. The initial block, also referred to as the header block, contains the hash value of the following block in the series, while the second block contains the IoT device's identification number, detected values, essential data, and its unique hash. If the second block is the final one, its hash value is not provided and is instead designated as either undefined or zero. The simulation used NS-3 for realistic scenario replication, and Ganache GUI with MetaMask was integrated into React.js for deploying the smart contract on a local Ethereum blockchain.

Figure 9 summarizes The Ganache graphical user interface, a decentralized Ethereum blockchain utilized in the study. It shows each account's digital address and an allocation of 100 ETH, with each account secured by a unique private key. Additionally, 16 blocks are created for various transactions conducted through the online interface. Figure 10 highlights the exact location within block 16 where the smart contract is deployed.

The data depicted in Fig. 11 illustrates that having only a small number of transactions per block could negatively affect the throughput. On the other hand, as the number of transactions per block increased to ten, the throughput significantly

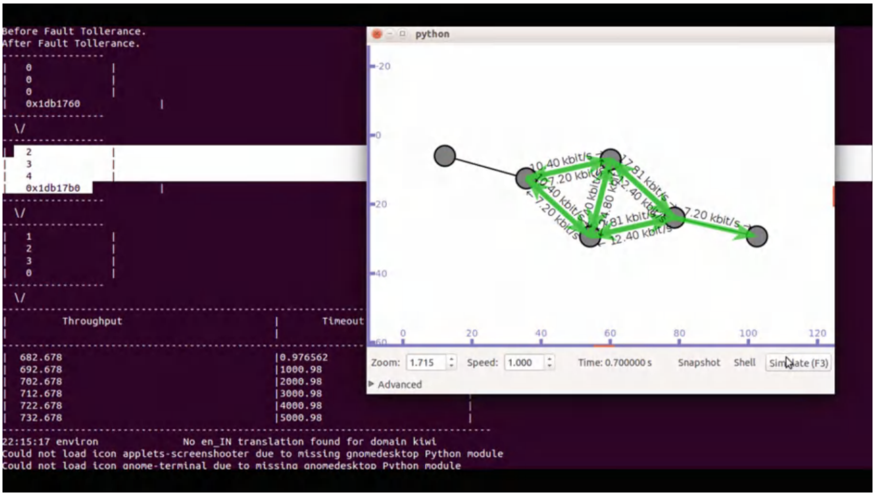


Fig. 8 The procedure of imitating and testing the ability to withstand faults

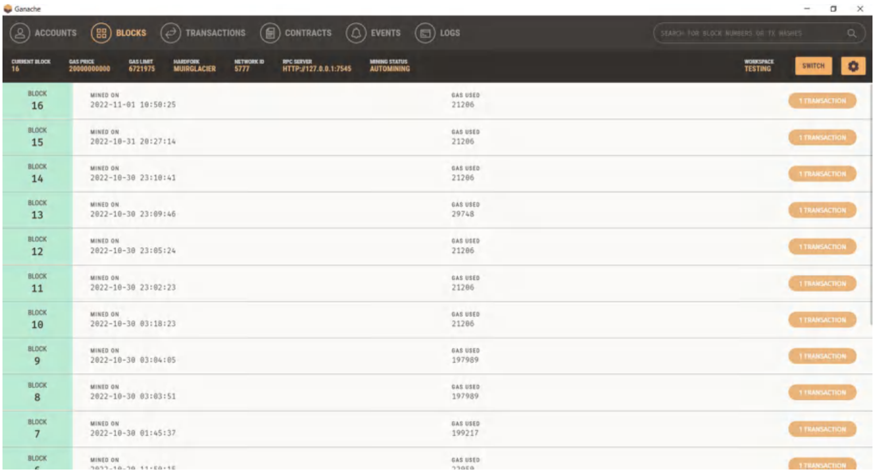


Fig. 9 Blocks generated

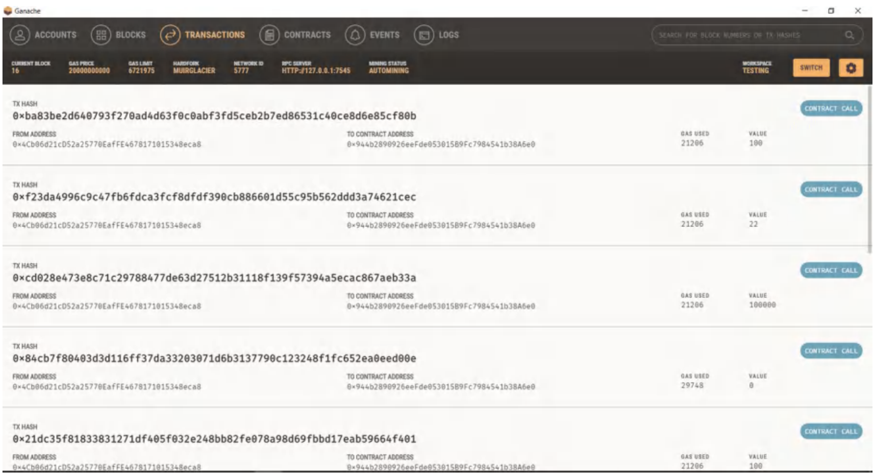


Fig. 10 Result about block 16 within the blockchain server

improved, but the speed advantage diminished. In this particular network configuration, a maximum throughput of 350 transactions per second was achieved with approximately 100 transactions per block.

Figure 12 the graph illustrates the changes in the size of the Blockchain's state database in megabytes, as displayed in Excel. The blue and orange colors depict the actual size (sstorage) and the size on disk (dsstorage) respectively. The information indicates that the state database's size consistently increased during multiple tests involving 5–600 product records. These findings align with expectations, considering the constant size of the product data utilized in the experiments.

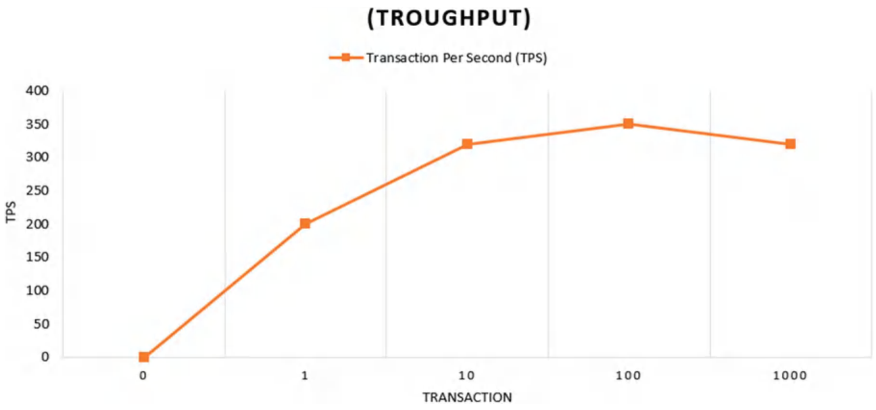


Fig. 11 Throughput

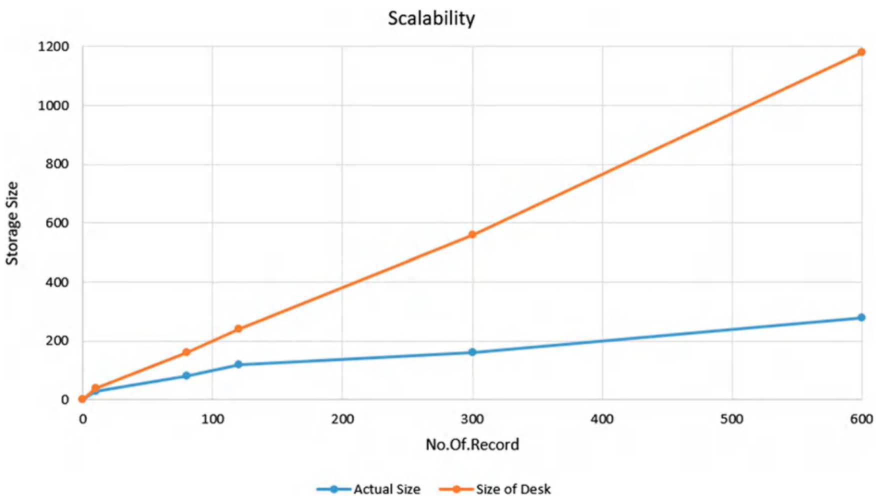


Fig. 12 Scalability

Figure 13 illustrates the difference in CPU and memory usage between Proof of Work (PoW) and Proof of Stake (PoS). PoW uses CPU and memory more intensively because of redundant mining operations and the computational burden of spreading blocks among nodes. This high demand for resources makes PoW unsuitable for IoT devices with their limited design capabilities.

In contrast, PoS requires much less computational power, leading to lower CPU and memory usage compared to PoW. PoS also decreases block confirmation times and energy consumption. As network traffic grows, PoW experiences scalability issues and packet delays, while PoS offers a more practical and efficient alternative with its lighter resource demands.

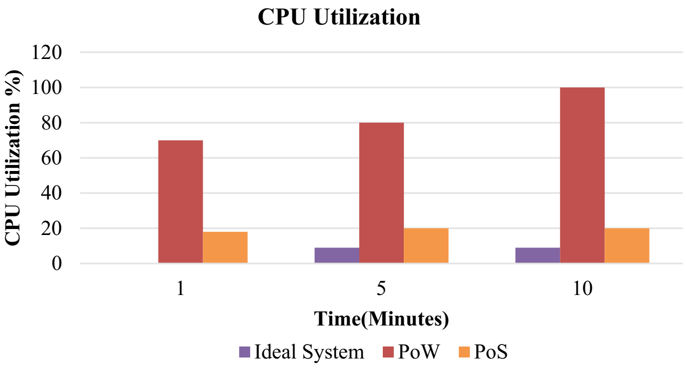


Fig. 13 CPU utilization

7 Conclusions

This paper introduced LightBlock, a scalable and lightweight blockchain model designed for IoT services in the public cloud, such as smart home management. It offers robust security and privacy, potentially enhancing IoT security for cloud-based services. LightBlock aims to support a national IoT infrastructure by ensuring end-to-end security while addressing IoT devices' limitations. The proposed design includes an adaptable hash function that improves blockchain network availability, performance, and efficiency, while minimizing space and power use on IoT devices. This lightweight approach enhances network scalability and is suitable for devices with latency under 1 s, optimizing utility and energy efficiency and reducing data redundancy.

References

1. P. Alvares, L. Silva, and N. Magaia, "Blockchain-based solutions for uav-assisted connected vehicle networks in smart cities: A review, open issues, and future perspectives," in *Telecom*, 2021, pp. 108–140.
2. A. Dhar Dwivedi, R. Singh, K. Kaushik, R. Rao Mukkamala, and W. S. Alnumay, "Blockchain and artificial intelligence for 5G-enabled Internet of Things: Challenges, opportunities, and solutions," *Transactions on Emerging Telecommunications Technologies,* vol. 35, p. e4329, 2024.
3. M. Iansiti and K. R. Lakhani, "The truth about blockchain," *Harvard business review,* vol. 95, pp. 118–127, 2017.
4. D. Burkhardt, M. Werling, and H. Lasi, "Distributed ledger," in *2018 IEEE international conference on engineering, technology and innovation (ICE/ITMC)*, 2018, pp. 1–9.
5. I. L. H. Alsammak, M. F. Alomari, I. S. Nasir, and W. H. Itwee, "A model for blockchain-based privacy-preserving for big data users on the internet of thing," *Indones. J. Electr. Eng. Comput. Sci,* vol. 26, pp. 974–988, 2022.
6. D. Yang, C. Long, H. Xu, and S. Peng, "A review on scalability of blockchain," in *Proceedings of the 2020 the 2nd International Conference on Blockchain Technology*, 2020, pp. 1–6.

7. E. P. Yadav, E. A. Mittal, and H. Yadav, "IoT: Challenges and issues in Indian perspective," in *2018 3rd International Conference On Internet of Things: Smart Innovation and Usages (IoT-SIU)*, 2018, pp. 1–5.
8. G. Papadodimas, G. Palaiokrasas, A. Litke, and T. Varvarigou, "Implementation of smart contracts for blockchain based IoT applications," in *2018 9th International Conference on the Network of the Future (NOF)*, 2018, pp. 60–67.
9. S. Saxena, B. Bhushan, and M. A. Ahad, "Blockchain based solutions to secure IoT: Background, integration trends and a way forward," *Journal of Network and Computer Applications,* vol. 181, p. 103050, 2021.
10. S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, and F.-Y. Wang, "Blockchain-enabled smart contracts: architecture, applications, and future trends," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 49, pp. 2266–2277, 2019.
11. F. Lombardi, L. Aniello, S. De Angelis, A. Margheri, and V. Sassone, "A blockchain-based infrastructure for reliable and cost-effective IoT-aided smart grids," 2018.
12. A. Dohr, R. Modre-Opsrian, M. Drobics, D. Hayn, and G. Schreier, "The internet of things for ambient assisted living," in *2010 seventh international conference on information technology: new generations*, 2010, pp. 804–809.
13. E. A. Wahdain, A. S. Baharudin, and M. N. Ahmad, "Big data analytics in the Malaysian public sector: the determinants of value creation," in *Recent Trends in Data Science and Soft Computing: Proceedings of the 3rd International Conference of Reliable Information and Communication Technology (IRICT 2018)*, 2019, pp. 139–150.
14. N. Giuffrida, J. Fajardo-Calderin, A. D. Masegosa, F. Werner, M. Steudter, and F. Pilla, "Optimization and machine learning applied to last-mile logistics: A review," *Sustainability,* vol. 14, p. 5329, 2022.
15. I. Sanchez, R. Satta, I. N. Fovino, G. Baldini, G. Steri, D. Shaw*, et al.*, "Privacy leakages in smart home wireless technologies," in *2014 international carnahan conference on security technology (ICCST)*, 2014, pp. 1–6.
16. C. Yang, B. Yuan, Y. Tian, Z. Feng, and W. Mao, "A smart home architecture based on resource name service," in *2014 IEEE 17th International Conference on Computational Science and Engineering*, 2014, pp. 1915–1920.
17. V. H. Bhide and S. Wagh, "i-learning IoT: An intelligent self learning system for home automation using IoT," in *2015 international conference on communications and signal processing (iccsp)*, 2015, pp. 1763–1767.
18. M. Gurunathan, M. A. Mahmoud, and M. N. Mahdi, "A systematic review on the challenges of adopting advanced security solutions onto IoT-based Smart Devices," *J. Theor. Appl. Inf. Technol,* vol. 100, pp. 1–17, 2022.
19. I. L. H. Alsammak, M. A. Mahmoud, H. Aris, M. AlKilabi, and M. N. Mahdi, "The use of swarms of unmanned aerial vehicles in mitigating area coverage challenges of forest-fire-extinguishing activities: A systematic literature review," *Forests,* vol. 13, p. 811, 2022.
20. Z. Lin, H. Niu, K. An, Y. Wang, G. Zheng, S. Chatzinotas*, et al.*, "Refracting RIS-aided hybrid satellite-terrestrial relay networks: Joint beamforming design and optimization," *IEEE Transactions on Aerospace and Electronic Systems,* vol. 58, pp. 3717–3724, 2022.
21. K. An, M. Lin, J. Ouyang, and W.-P. Zhu, "Secure transmission in cognitive satellite terrestrial networks," *IEEE Journal on Selected Areas in Communications,* vol. 34, pp. 3025–3037, 2016.
22. I. L. H. Alsammak, M. A. Mahmoud, S. S. Gunasekaran, A. N. Ahmed, and M. AlKilabi, "Nature-inspired drone swarming for wildfires suppression considering distributed fire spots and energy consumption," *IEEE Access,* 2023.
23. M. A. Mahmoud, M. Gurunathan, R. Ramli, K. A. Babatunde, and F. H. Faisal, "Review and Development of a Scalable Lightweight Blockchain Integrated Model (LightBlock) for IoT Applications," *Electronics,* vol. 12, p. 1025, 2023.

Hybrid Deep Learning for Detecting Hate Speech Across Social Media Platforms

Loke Kevin, Chai Ting Yee, Chia Zhi Liang, Stephen Yong, Yeap Cheng Han, Muaadh Mukred, and Fathey Mohammed

Abstract The proliferation of social media platforms has enhanced global communication and facilitated the spread of hate speech, posing a significant challenge to online communities. This study investigates developing and implementing hybrid deep learning models for automatically detecting hate speech across multiple social media platforms. The research aims to improve detection accuracy and efficiency by combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks. The study leverages a dataset of approximately 50,000 comments from the "Piers Morgan Uncensored" YouTube channel. Significant improvements in hate speech detection were observed, with the CNN model achieving an accuracy of 99%, representing a substantial improvement over traditional methods. Practical implications include enhanced content moderation capabilities, allowing online platforms to more effectively identify and mitigate hate speech, reducing the burden on human moderators, and thus fostering a safer and more inclusive digital environment.

Keywords Hate speech detection · Deep learning · Hybrid models · Sentiment analysis · Social media

1 Introduction

Social media has revolutionized communication, enabling instantaneous interaction and information exchange across the globe. Platforms like Twitter, Facebook, and YouTube have become integral to daily life, offering forums for public discourse, personal expression, and community building. However, this digital transformation has also given rise to significant challenges, one of the most pressing being the proliferation of hate speech [1]. Defined as a language that demeans, insults, or threatens individuals based on attributes such as race, religion, ethnicity, sexual orientation, or

L. Kevin · C. T. Yee · C. Z. Liang · S. Yong · Y. C. Han · M. Mukred (✉) · F. Mohammed
Department of Business Analytics, Sunway Business School, Sunway University, 5, Jalan University, 47500 Petaling Jaya, Bandar Sunway, Selangor, Malaysia
e-mail: muaadhm@sunway.edu.my

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_16

gender, hate speech poses a severe threat to the fabric of online communities. It fosters divisiveness and animosity and can incite real-world violence and discrimination [2].

Addressing hate speech is complex due to its varied forms and the subtlety of language. Simple keyword-based filters are insufficient, as hate speech often employs coded language, euphemisms, and context-dependent nuances. This necessitates more sophisticated approaches capable of understanding the intricacies of human language and the context in which it is used [3, 4]. In this regard, deep learning models, particularly those leveraging neural network architectures, offer promising solutions. These models excel in recognizing patterns and making predictions based on large datasets, making them suitable for tasks like sentiment analysis and hate speech detection [4].

This research focuses on developing hybrid deep learning models that combine Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks to enhance the detection of hate speech on social media platforms. By integrating these models with sentiment analysis tools such as Valence Aware Dictionary and sEntiment Reasoner (VADER) and TextBlob, the study aims to achieve a more nuanced understanding and accurate classification of hate speech. The chosen dataset, comprising approximately 50,000 comments from the “Piers Morgan Uncensored” YouTube channel, provides a rich source of data for analysis, reflecting diverse user-generated content and potential instances of hate speech.

Despite the progress in hate speech detection, a significant gap remains in achieving high accuracy and efficiency across diverse social media platforms [5]. Existing models often struggle with the complexity and variability of language used in hate speech [5, 6]. This study addresses this gap by developing a hybrid deep learning model that leverages the strengths of CNNs and LSTMs combined with advanced sentiment analysis tools. The primary contributions of this research include demonstrating a significant improvement in detection accuracy, with the CNN model achieving 99% accuracy, providing a scalable and efficient solution for hate speech detection applicable across multiple social media platforms, and enhancing the practical implications for content moderation by offering automated preliminary filtering of harmful language and supporting human moderators in maintaining safer online environments. This paper outlines the methodology employed, discusses the results and their implications, and highlights the contributions and future directions for research in this critical area.

2 Literature Review

The literature on hate speech detection and mitigation on social media is vast and multifaceted, reflecting this issue’s complexity and critical importance. Hate speech, defined as any communication that belittles or discriminates against individuals based on inherent characteristics, is a pervasive problem in digital communication [7, 8]. Traditional approaches, such as keyword filtering, often fail due to hate speech’s dynamic and context-sensitive nature [2]. Therefore, advanced methods leveraging machine learning (ML) and deep learning (DL) are increasingly adopted to address this issue.

Early ML techniques, such as support vector machines (SVM), have been employed to classify text into hate speech and non-hate speech categories [9]. Tiwari et al. [10] demonstrated the effectiveness of SVM in sentence-level analysis, outperforming other classifiers like Random Forest and Decision Trees in certain scenarios. However, these traditional ML approaches often struggle with complex language patterns and contextual nuances, necessitating more sophisticated solutions [11].

Deep learning models, particularly CNNs and LSTMs, have shown significant promise in hate speech detection [12, 13]. CNNs are adept at capturing local features and spatial hierarchies in text data, while LSTMs excel at handling sequential data and maintaining context over longer text spans [14]. Abdullah [15] highlighted the efficacy of a CNN-LSTM network in sentiment and emotion detection, showcasing the potential of hybrid models in understanding complex linguistic structures.

Multimodal approaches, incorporating text, audio, and video analysis, have also been explored to capture the full spectrum of online interactions [16]. Perifanos and Goutsos [17] demonstrated that integrating multiple data modalities significantly improves the detection of hate speech in multimedia content, outperforming traditional text-based methods. This highlights the importance of considering various forms of digital content to develop comprehensive detection systems.

Hybrid models combining different DL architectures have been proposed to leverage their respective strengths. Simon et al. [18] developed a hybrid model integrating CNNs and LSTMs with GloVe embeddings, achieving enhanced accuracy and efficiency in hate speech detection. Such models are particularly effective in handling the multifaceted nature of hate speech, which often involves intricate language patterns and context-dependent meanings.

The challenge of detecting hate speech is further compounded by the need to account for linguistic diversity and cultural nuances. Mahajan et al. [19] emphasized the difficulty of distinguishing hate speech from vulgar language, underscoring the importance of context and subtlety in language. Similarly, Nasir et al. [20] explored hate speech detection in Urdu, demonstrating the efficacy of transfer learning in low-resource languages and highlighting the need for adaptable models.

Ethical considerations are paramount in the deployment of Artificial Intelligence (AI) and ML technologies for hate speech detection. Cortiz and Zubiaga [21] discussed the ethical implications of using AI to establish causal links between media remarks and violent actions, advocating for balanced approaches that respect user privacy and freedom of expression. Solovev and Pröllochs [22] examined the use of moralized language to predict hate speech, suggesting that nuanced language understanding is critical for effective detection.

This study builds on these findings by developing a hybrid deep learning model that integrates CNNs, LSTMs, and sentiment analysis tools to enhance the detection of hate speech on social media platforms. The importance of context dependence could address one of the core challenges of reliably detecting hate speech. Integrating sentiment analysis could build a new lexicon for hate speech detection, providing a scalable and efficient solution for hate speech detection applicable across multiple social media platforms, and enhancing the practical implications for content moderation by offering automated preliminary filtering of harmful language and supporting human moderators in maintaining safer online environments.

The literature underscores the need for diverse and robust models capable of capturing the intricacies of human language and context. This study addresses these needs by developing a hybrid deep learning model that integrates CNNs, LSTMs, and advanced sentiment analysis tools like VADER and TextBlob. The primary contribution is the emphasis on the importance of context through sentiment analysis, which helps eliminate one of the core complexities of detecting hate speech. Unlike previous hybrid models, this approach integrates sentiment analysis to refine classification, providing a nuanced understanding of sentiment and improving performance in distinguishing hate speech. The model is scalable and efficient, applicable across multiple social media platforms, and achieves a significant improvement in detection accuracy, reaching 99%. It also adapts to new forms of hate speech over time and provides automated preliminary filtering, reducing the burden on human moderators and supporting safer online environments.

3 Methodology

This study employs a systematic approach to develop and implement hybrid deep learning models for the automatic detection of hate speech on social media platforms. The methodology encompasses data collection, preprocessing, model development, and evaluation.

3.1 Data Collection

The dataset consists of approximately 50,000 comments from the "Piers Morgan Uncensored" YouTube channel, focusing on reactions to interviews with Andrew Tate and Piers Morgan. The comments were collected using the YouTube Data API, ensuring a comprehensive representation of user sentiments. The data collection period spanned twenty-six days, capturing a diverse range of user interactions and potential instances of hate speech.

3.2 Data Preprocessing

Data preprocessing is critical for ensuring the quality and relevance of the dataset [23, 24]. The steps include text cleaning, tokenization, normalization, stop words removal, lemmatization, and sentiment labeling [24]. Text cleaning involves removing URLs, special symbols, and excessive whitespace, while retaining emojis and punctuation for sentiment analysis. Tokenization converts text into sequences of words, and normalization ensures uniformity by converting text to lowercase. Stop words removal eliminated words that do not contribute to sentiment analysis. Lemmatization reduced complex words into their base form [25]. Sentiment labeling is performed using VADER and TextBlob to classify comments into positive, negative, or neutral categories.

3.3 Model Development

The hybrid deep learning model combines CNNs and LSTMs. CNNs are used to capture spatial hierarchies and patterns within the text data, while LSTMs handle sequential data and maintain context over longer text spans. The model architecture integrates the outputs of CNN layers with LSTM layers, creating a robust framework for detecting complex language patterns associated with hate speech. The model was implemented using TensorFlow and Keras libraries.

3.4 Model Training and Evaluation

The dataset was split into training, validation, and test sets. The hybrid model was compiled with binary cross-entropy as the loss function and the Adam optimizer with a learning rate of 0.0005. Training was conducted over 30 epochs with early stopping criteria to prevent overfitting. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1-score.

4 Results and Discussion

4.1 Data Analysis

The initial analysis using VADER and TextBlob revealed significant variations in sentiment across the dataset. VADER, specifically designed for social media text, demonstrated superior performance in capturing nuanced sentiments indicative of hate speech. This initial sentiment classification laid the groundwork for the subsequent application of deep learning models.

Textblob

In the preliminary phase of employing TextBlob for sentiment analysis, each comment within the dataset underwent a meticulous process of sentiment polarity calculation. This involved the assignment of a numerical score to each comment, indicative of its overall sentiment orientation. Subsequently, leveraging these polarity scores, a classification scheme was implemented to categorize comments into distinct sentiment categories: “positive,” “negative,” and “neutral”. Following the classification process, a comprehensive analysis of the sentiment distribution within the dataset was conducted using histograms. These visual representations provided a clear overview of the distribution of sentiment categories, allowing for an insightful exploration of the prevalence of positive, negative, and neutral sentiments across the dataset.

Figure 1 revealed compelling insights into the sentiment landscape of the dataset. Specifically, it unveiled that a significant proportion of comments exhibited a positive sentiment orientation, with a total count of 20,069 comments classified as “positive.” Additionally, the analysis highlighted a considerable presence of comments categorized as “neutral,” with a total count of 19,168 comments falling into this category. Conversely, the number of comments classified as “negative” was comparatively lower, with a total count of 10,946 comments expressing predominantly negative sentiment.

VADER

In the subsequent phase of the analysis, VADER (Valence Aware Dictionary and Sentiment Reasoner) was utilized for sentiment analysis. To facilitate this process, the Sentiment Intensity Analyzer library was imported into Python, providing access to the VADER sentiment analysis tool. The VADER sentiment analysis tool employs

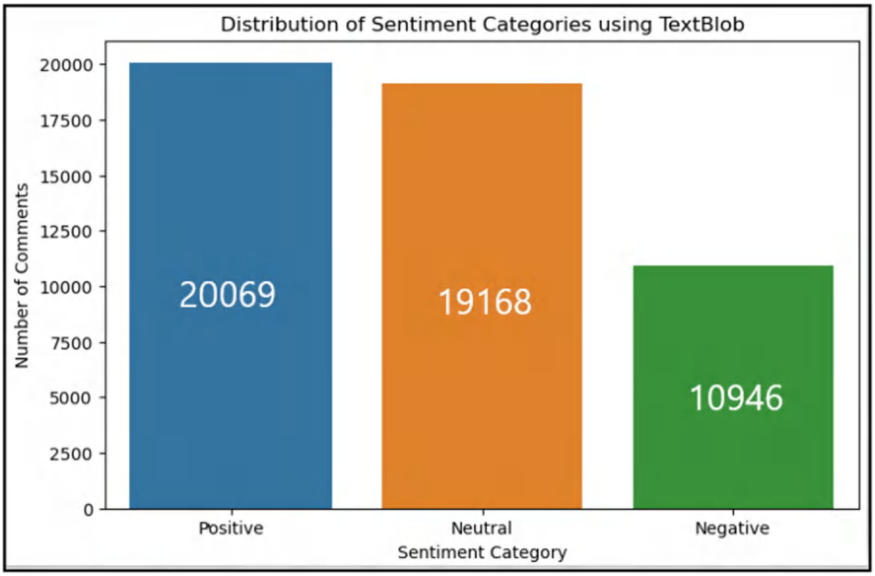


Fig. 1 Distribution of sentiment analysis categories using Textblob

a sophisticated algorithm to assess the sentiment expressed within textual data, offering four distinct sentiment intensity measures: "positive," "negative," "neutral," and "polarity." Each comment within the dataset underwent evaluation using these sentiment intensity measures, resulting in the calculation of corresponding scores for each sentiment category.

Following the computation of sentiment intensity scores, comments were categorized into discrete sentiment categories based on the polarity score. Notably, the sentiment polarity ranges from 0.0 to 0.2 exhibits the highest frequency, approximately 14,000. The initial day recorded the highest peak for the average sentiment score, reaching 0.08. Following this peak, there has been a subsequent fluctuation with some slight increases and decreases. Upon completion of the sentiment classification process, a detailed analysis of the sentiment distribution within the dataset was conducted using histograms. Figure 2 provided a visual depiction of the prevalence of positive, negative, and neutral sentiments across the dataset, facilitating a nuanced exploration of sentiment dynamics.

Figure 3 analysis yielded insightful findings regarding the sentiment distribution within the dataset. Notably, a substantial number of comments were classified as "positive," with a total count of 19,832 comments exhibiting predominantly positive sentiment. Additionally, the analysis revealed a significant presence of comments categorized as "negative," totaling 17,807 comments expressing predominantly negative sentiment. Furthermore, a considerable proportion of comments were classified as "neutral," with a total count of 12,544 comments demonstrating a lack of discernible sentiment.

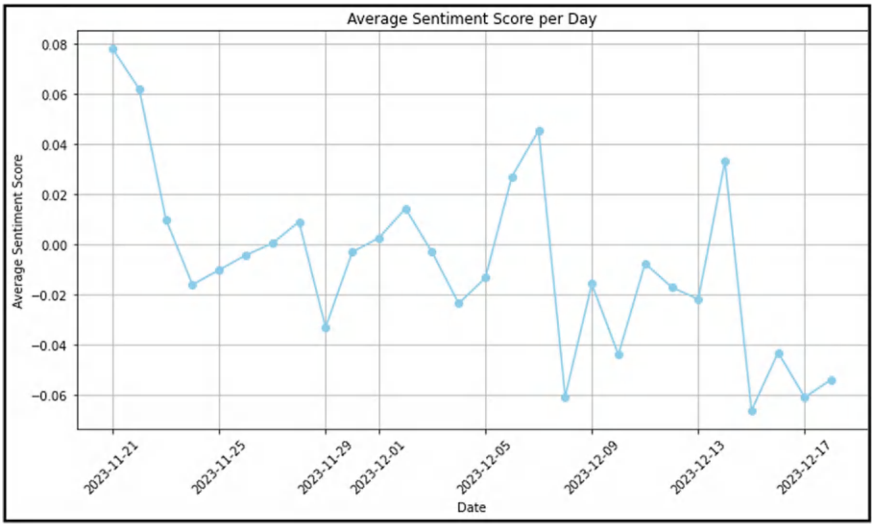


Fig. 2 Average sentiment score per day over the 26 days period

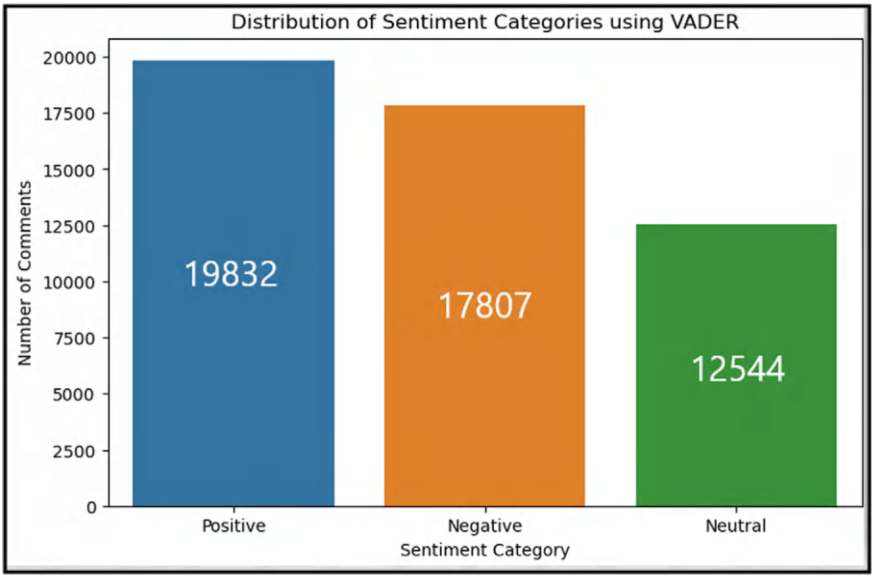


Fig. 3 Distribution of sentiment analysis categories using VADER

4.2 *Model Development*

After performing sentiment analysis of the dataset using both TextBlob and VADER, the result output underwent human evaluation to determine its accuracy specifically on the labeled negative sentiment. Between both TextBlob and VADER, VADER classified negative sentiment more accurately compared to TextBlob. Therefore, hate speech was labeled in the dataset by any score that is below -0.50. Human evaluation was utilized to ensure false positive and false negative is minimized and to arbiter sarcasm heavy comments into their appropriate category.

Label Encoding

After sentiment analysis, before fitting the data into the machine learning models, label encoding is conducted on the 'sentiment' column to convert the categorical labels into numeral labels. Label encoding is chosen here for its simplicity and suitability for ordinal categorical data like sentiment labels ('negative', 'neutral', 'positive'). Since sentiment labels have an inherent order (negative < neutral < positive), assigning numerical values (0, 1, 2) preserves this order, making label encoding an appropriate choice. Other encoding methods like one-hot encoding could also be used, but they would result in creating additional binary columns for each category, which might not be necessary for ordinal data. Additionally, label encoding saves memory and computational resources compared to one-hot encoding, especially when dealing with a large number of categories. Therefore, label encoding is preferred in this context for its efficiency and preservation of ordinal information.

Under-Sampling

Moving on, under-sampling is performed on the sentiment labels after label encoding. Under-sampling is a technique used to address class imbalance in datasets, specifically targeting sentiment data. Class imbalance occurs when one class (e.g., 'negative', 'neutral', 'positive' sentiments) significantly outnumbers the others, potentially leading to biased model predictions. In this context, under-sampling ensures that each sentiment class is represented equally by randomly removing excess samples from overrepresented classes until the dataset achieves balanced class distribution. This approach is crucial for training machine learning models, as it prevents the model from favoring the majority class and allows it to learn from all classes equally. By implementing under-sampling, the code promotes fair and unbiased model training, thereby enhancing the model's ability to accurately predict sentiment across all classes.

Count Unique Words

Before implementing the deep learning model, the unique words in the text are counted as counting unique words in text data is fundamental for deep learning tasks involving RNN and CNN. Each unique word represents a feature, and its frequency or presence in a document is crucial for understanding its context. This process enables the creation of a vocabulary, serving as the foundation for converting text data into numerical form, essential for model processing. The size of the vocabulary impacts the model's complexity and computational requirements. A larger vocabulary increases the dimensionality of the input space, which can make the model more complex and computationally intensive [26]. By counting unique words, we gain insights into the vocabulary size, allowing us to optimize model architecture and resource allocation accordingly. By analyzing the distribution of unique words, valuable insights can be gained into the richness and diversity of the text data, influencing preprocessing strategies and model design. Ultimately, counting unique words plays a pivotal role in preparing data for deep learning models, ensuring meaningful input representation and enhancing the model's performance in processing textual information.

Train-Test Split

Firstly, the target variable 'y' is set as the text and the independent variable 'x' is set as the sentiment of the text. Moving on, implementing a train-test split is an essential practice in machine learning model development to ensure robust performance assessment. This process involves dividing the dataset into two subsets: one for training the model and another for evaluating its performance. By partitioning the data in this manner, the model's ability to generalize to unseen data can be effectively tested. Typically, this split allocates a majority of the data to the training set, enabling the model to learn patterns and relationships present in the dataset. Meanwhile, the testing set, comprising a smaller portion of the data, remains untouched during model training and serves as an independent sample for evaluating the model's performance. This separation mitigates the risk of overfitting, where the model performs well on the training data but fails to generalize to new, unseen data, a common challenge in machine learning. In the dataset, the dataset is partitioned into 80% training and 20% testing split.

Additionally, the random seed is set as 42, which ensures reproducibility, as it fixes the randomization process, leading to consistent results across multiple runs. After the split, the shape of the dataset is checked across all test and train data to ensure the split is done correctly.

Tokenizer

In order to perform the tokenization, the tokenizer class from TensorFlow's Keras API is used as it plays a pivotal role in preprocessing text data for natural language processing tasks. By tokenizing text, the tokenizer breaks down input sentences or documents into individual words or tokens, enabling numerical representation and subsequent processing by machine learning models. The tokenizer is employed to convert textual data from both the training and testing sets into sequences of integers, where each integer corresponds to a unique word in the vocabulary [27]. This process facilitates the creation of a standardized input format for deep learning models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), which typically operate on numerical data [28, 29]. Additionally, the use of padding with the pad_sequences function ensures that all sequences have the same length, a requirement for input into neural networks. Overall, the Tokenizer enables efficient text preprocessing, transforming raw text data into a format suitable for training and evaluating deep learning models, thereby enhancing their performance in tasks such as sentiment analysis, text classification, and more.

Figure 4, displays a line graph tracking the daily counts of toxic (hate speech) and positive comments over a period of time. The graph shows a significant peak in positive comments on November 23, 2021, with the count reaching a maximum of 3,555, on the same day toxic comments also hit their highest at 1,029. Following this peak, the number of both comment types declines sharply, with positive comments reducing at a faster rate than toxic comments. Over the subsequent days, the number of toxic comments levels off, maintaining a steady, low frequency, while positive comments also stabilize but at a slightly higher frequency than toxic comments. This visualization highlights a particular event or series of events that caused a surge in both positive and toxic online discourse on a specific date, followed by a return to lower levels of activity.

4.3 Model Performance

This section is devoted to present the findings of the Model performance as follows.

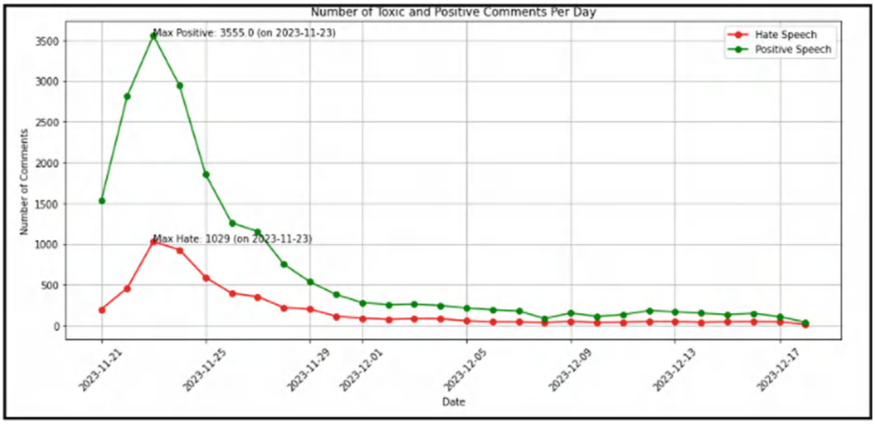


Fig. 4 Number of toxic and positive comments per day over the 26 days period

SVM Performance

The SVM model achieved an accuracy of 0.96, with precision, recall, and F1-score of 0.98, 0.95, and 0.96, respectively. Despite its robustness, SVM was outperformed by the neural network models in handling the complexities of the dataset.

CNN Performance

The CNN model exhibited the best overall performance with an accuracy of 0.99, precision of 0.99, recall of 0.98, and an F1-score of 0.99. CNNs effectively captured local patterns and spatial hierarchies in the text, making them well-suited for detecting hate speech.

RNN Performance

The RNN model, with its ability to handle sequential data, achieved an accuracy of 0.98, precision of 1.0, recall of 0.96, and an F1-score of 0.98. The RNN's performance, particularly its perfect precision score, indicates its effectiveness in predicting true positive instances without misclassifying non-hate speech content.

4.4 Comparative Analysis

The comparative analysis highlights the strengths of each model. While SVMs are known for their robustness in high-dimensional spaces, they were outperformed by CNNs and RNNs, which better captured the complexities of social media text. CNN's superior accuracy and recall demonstrate its proficiency in recognizing hate speech patterns. Meanwhile, the RNN's exceptional precision underscores its capability in handling sequential dependencies and temporal context in text data. Table 1 summarizes the evaluation metrics comparison for SVM, CNN and RNN.

Table 1 Comparison of accuracy, precision, recall, and F1-score among machine learning models

Model	Accuracy	Precision	Recall	F-1 score
SVM	0.96	0.98	0.95	0.96
CNN	0.99	0.99	0.98	0.99
RNN	0.98	1.0	0.96	0.98

Table 1 summarizes the performance metrics of different machine learning models used in the study. Accuracy refers to the proportion of true results among the total number of cases examined. Precision is the ratio of correctly predicted positive observations to the total predicted positives. Recall (Sensitivity) is the ratio of correctly predicted positive observations to all observations in actual class. F1-Score is the weighted average of Precision and Recall.

4.5 Sentiment Dynamics and Hate Speech Detection

The detailed analysis of sentiment dynamics revealed that a significant portion of comments exhibited negative sentiments. VADER identified 17,807 comments as negative, reflecting the prevalence of hostile language within the dataset. The use of human evaluation to refine these classifications further ensured the accuracy of the hate speech labels. This process involved reviewing the outputs from VADER and TextBlob to minimize false positives and negatives, particularly in detecting sarcasm and nuanced expressions of negativity.

4.6 Temporal Analysis

A temporal analysis of sentiment scores indicated that peaks in negative sentiment often coincided with specific events or discussions within the dataset. For example, a notable peak in negative comments was observed on November 23, 2021, suggesting a significant event or controversy related to the interviews. This temporal pattern underscores the importance of contextualizing hate speech within specific time frames to better understand its triggers and dynamics.

4.7 Visualization and Insights

Visualizations, including histograms and pie charts, provided a clear depiction of sentiment distribution and trends. Histograms showed the prevalence of positive, neutral, and negative sentiments, while pie charts illustrated the proportion of each sentiment category. These visual tools facilitated a nuanced exploration of sentiment dynamics, revealing that a substantial number of comments were classified as positive (19,832) and neutral (12,544), alongside the significant presence of negative comments.

4.8 Model Robustness and Scalability

The robustness and scalability of the models were evaluated through various metrics and visualizations. The training loss and validation loss for both CNN and RNN models exhibited a consistent downward trend, indicating effective learning and minimal overfitting. Additionally, the Receiver Operating Characteristic (ROC) curves for all models demonstrated high Area Under the Curve (AUC) values, further validating their performance.

The training time analysis revealed that while SVMs are computationally efficient, requiring only 6.91 s, the deeper architectures of RNNs and CNNs necessitated longer training durations (114.88 s and 76.29 s, respectively). This trade-off between computational efficiency and model complexity highlights the need for balanced approaches depending on the specific requirements of the application.

5 Contributions

This research makes significant contributions to the field of hate speech detection on social media. By integrating CNNs and LSTMs within a hybrid model framework, this study enhances the accuracy and efficiency of hate speech detection algorithms. The use of sentiment analysis tools like VADER and TextBlob further refines the classification process, ensuring more reliable detection of nuanced expressions of hate speech.

The use of sentiment analysis in the hybrid approach allows the model to better understand the context and emotional tone of the content, enhancing its ability to detect subtle and implicit hate speech that might be missed by traditional keyword-based approaches. This context-based understanding is critical in identifying and mitigating evolving language patterns and emerging threats.

The study also addresses the ethical implications of deploying AI for content moderation, advocating for balanced approaches that protect user privacy while mitigating online abuse.

5.1 Practical Implications

The practical implications of this study are significant for both academia and industry. The development of hybrid deep learning models capable of accurately detecting hate speech can enhance content moderation systems, making online platforms safer and more inclusive. These models can assist human moderators by reducing their exposure to potentially traumatizing content and providing automated preliminary filtering of harmful language.

Moreover, the insights gained from this study can inform the development of policies and guidelines for online conduct, balancing the protection of free expression with the need to curb toxicity. The findings also underscore the importance of continuous monitoring and updating of hate speech detection systems to adapt to evolving language patterns and emerging threats.

5.2 Real-World Applications and Challenges in Deployment

In practical terms, deploying these hybrid deep learning models can significantly improve the efficiency and effectiveness of content moderation on various social media platforms and other online communities. By automating the initial screening process, these models can handle large volumes of data efficiently, allowing human moderators to focus on more complex cases. This not only speeds up the moderation process but also reduces the psychological burden on moderators who would otherwise be exposed to harmful content regularly.

However, deploying such systems comes with challenges. Ensuring the models are robust and unbiased is critical to avoid over-censorship or the inadvertent silencing of marginalized voices. Additionally, the dynamic nature of language, especially in the context of slang and coded language used to evade detection, requires continuous updates and retraining of models. There's also the technical challenge of integrating these models seamlessly into existing moderation frameworks across multitude of platforms and ensuring they operate in real-time without significant latency.

6 Future Works

Future research should focus on expanding the dataset to include diverse social media platforms and languages, capturing a broader spectrum of online discourse. Additionally, the development and implementation of more advanced hybrid models, including semi-supervised and unsupervised learning techniques, can further enhance detection accuracy by utilizing both labeled and unlabeled data. Ethical considerations remain paramount, necessitating transparent frameworks for user consent and bias mitigation. Long-term impact assessment should focus on developing a new lexicon to keep pace with evolving language and slang in online discourse. Context-aware models or tools are essential for accurate detection, as they can better understand the nuances of changing language patterns. Interdisciplinary collaboration with social media companies, policymakers, and interdisciplinary research teams will enrich methodologies and foster holistic solutions to complex societal challenges, ensuring practical scalable solutions.

7 Conclusion

This study demonstrates the efficacy of hybrid deep learning models in detecting hate speech on social media platforms. By leveraging the strengths of CNNs and LSTMs and integrating sentiment analysis tools, the models developed in this research exhibit high accuracy and efficiency. The findings highlight the importance of nuanced language understanding and contextual analysis in identifying hate speech. The practical implications for content moderation and policy development are significant, contributing to safer and more inclusive online environments. Continued research and development in this area are essential to address the evolving challenges of online hate speech and ensure the responsible deployment of AI technologies.

References

1. Tariq, R., *Hate speech or healthy speech, user friendly or user anxiety: An analysis of cyber world speech discourse.* Journal of Media Studies, 2019. **31**(1).
2. MacAvaney, S., et al., *Hate speech detection: Challenges and solutions.* PloS one, 2019. **14**(8): p. e0221152.
3. Abdelsamie, M.M., S.S. Azab, and H.A. Hefny, *A comprehensive review on Arabic offensive language and hate speech detection on social media: methods, challenges and solutions.* Social Network Analysis and Mining, 2024. **14**(1): p. 1–49.
4. Gupta, P. and S. Gupta, *Hate speech detection using machine learning models*, in *Computational Intelligence in the Industry 4.0.* 2024, CRC Press. p. 189–218.
5. Yin, W. and A. Zubiaga, *Towards generalisable hate speech detection: a review on obstacles and solutions.* PeerJ Computer Science, 2021. **7**: p. e598.
6. Nirmal, A., *Interpretable hate speech detection via large language model-extracted rationales.* 2024, Arizona State University.
7. Keighley, R., *Towards a Better Understanding of Hate*, in *Researching Hate as an Activist: Exploring LGBTQ+ Online Hate, Its Impacts, and Our Responsibility Towards Equality*. 2024, Springer. p. 85–102.
8. Kwarteng, J., *Understanding Misogynoir Online: Challenges In Identifying Intersectional Hate*. 2024, The Open University.
9. Subramanian, M., et al., *A survey on hate speech detection and sentiment analysis using machine learning and deep learning models.* Alexandria Engineering Journal, 2023. **80**: p. 110–121.
10. Tiwari, S., et al. *Social media sentiment analysis on Twitter datasets.* in *2020 6th international conference on advanced computing and communication systems (ICACCS).* 2020. IEEE.

11. Khan, W., et al., *Exploring the frontiers of deep learning and natural language processing: A comprehensive overview of key challenges and emerging trends.* Natural Language Processing Journal, 2023: p. 100026.
12. Revathi, S., et al. *Hate Speech Detection Using Deep Learning Algorithms.* in *International Conference on Intelligent Systems Design and Applications*. 2023. Springer.
13. Alhazmi, A., et al., *A systematic literature review of hate speech identification on Arabic Twitter data: research challenges and future directions.* PeerJ Computer Science, 2024. **10**: p. e1966.
14. Huang, F., et al., *Attention-emotion-enhanced convolutional LSTM for sentiment analysis.* IEEE transactions on neural networks and learning systems, 2021. **33**(9): p. 4332–4345.
15. Abdullah, M.A.-G., *Deep Learning for Sentiment and Emotion Detection in Multilingual Contexts.* 2018, The University of North Carolina at Charlotte.
16. Shoumy, N.J., et al., *Multimodal big data affective analytics: A comprehensive survey using text, audio, visual and physiological signals.* Journal of Network and Computer Applications, 2020. **149**: p. 102447.
17. Perifanos, K. and D. Goutsos, *Multimodal hate speech detection in Greek social media.* Multimodal Technologies and Interaction, 2021. **5**(7): p. 34.
18. Simon, Y., B.Y. Baha, and E.J. Garba, *A Multi-Platform Approach Using Hybrid Deep Learning Models for Automatic Detection of Hate Speech on Social Media.* Bima Journal of Science and Technology (2536–6041), 2022. **6**(02): p. 77–90.
19. Mahajan, E., H. Mahajan, and S. Kumar, *EnsMulHateCyb: Multilingual hate speech and cyberbully detection in online social media.* Expert Systems with Applications, 2024. **236**: p. 121228.
20. Nasir, S., A. Seerat, and M. Wasim. *Hate Speech Detection in Roman Urdu using Machine Learning Techniques.* in *2024 5th International Conference on Advancements in Computational Sciences (ICACS).* 2024. IEEE.
21. Cortiz, D. and A. Zubiaga, *Ethical and technical challenges of AI in tackling hate speech.* The International Review of Information Ethics, 2020. **29**.
22. Solovev, K. and N. Pröllochs, *Moralized language predicts hate speech on social media.* PNAS nexus, 2023. **2**(1): p. pgac281.
23. Bhandari, K., K. Kumar, and A.L. Sangal, *Data quality issues in software fault prediction: a systematic literature review.* Artificial Intelligence Review, 2023. **56**(8): p. 7839–7908.
24. Chi, W.W., et al., *Data Augmentation With Semantic Enrichment for Deep Learning Invoice Text Classification.* IEEE Access, 2024.
25. Khyani, D., et al., *An interpretation of lemmatization and stemming in natural language processing.* Journal of University of Shanghai for Science and Technology, 2021. **22**(10): p. 350–357.
26. Woensdregt, M.S., et al. *Why is scaling up models of language evolution hard?* in *Proceedings of the annual meeting of the cognitive science society*. 2021.
27. Pai, A., *What is tokenization in nlp? here's all you need to know*. 2020, Retrieved from Analytics Vidhya: https://www.analyticsvidhya.com/blog/2020….
28. Erkan, A. and T. Güngör, *Analysis of Deep Learning Model Combinations and Tokenization Approaches in Sentiment Classification.* IEEE Access, 2023. **11**: p. 134951–134968.
29. Korvigo, I., et al., *Putting hands to rest: efficient deep CNN-RNN architecture for chemical named entity recognition with no hand-crafted rules.* Journal of cheminformatics, 2018. **10**: p. 1–10.

AI-Driven Smart City Security and Surveillance System: A Bibliometric Analysis

Wong Qi Fan, Aalima Shaura Ismail, Fathey Mohammed, and Muaadh Mukred

Abstract Integrating advanced technologies such as artificial intelligence (AI) has become crucial for enhancing urban management and security as urban centers evolve into smart cities. This study conducts a bibliometric analysis to explore smart cities' security and surveillance dimensions, focusing on data protection and ethical AI use. By reviewing 745 articles indexed in Scopus from 1977 to 2023, we employ co-citation analysis, co-occurrence analysis, and bibliographic coupling to identify key thematic clusters and influential publications. The findings reveal the complex interplay between technological advancements and privacy concerns, highlighting the importance of a balanced approach to AI integration. This research provides valuable insights for urban planners and policymakers to develop strategies that enhance urban safety and efficiency while respecting individual privacy and fostering public trust. The study underscores the necessity of ethical standards and robust governance frameworks in deploying AI-powered security systems in smart cities.

Keywords Smart cities · Artificial intelligence (AI) · Security and surveillance · Bibliometric analysis

1 Introduction

The rapid advancement of technology has significantly transformed urban landscapes, giving rise to the concept of smart cities. A smart city integrates information and communication technologies (ICTs) and the Internet of Things (IoT) to manage a city's assets, including transportation systems, power plants, water supply networks, and waste management systems [1]. Smart cities use technology and data to improve urban life, enhance public services, and promote sustainability. Rapid urbanization

W. Q. Fan · A. S. Ismail · F. Mohammed (✉) · M. Mukred (✉)
Department of Business Analytics, Sunway Business School, Sunway University, 5, Jalan University, 47500 Petaling Jaya, Bandar Sunway, Selangor, Malaysia
e-mail: fatheym@sunway.edu.my

M. Mukred
e-mail: muaadhm@sunway.edu.my

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_17

and technological advancements like the Internet of Things (IoT) and big data are driving this evolution. Among these technological advancements, artificial intelligence (AI) has emerged as a cornerstone for enhancing urban management and security. Artificial Intelligence (AI) plays a crucial role by analyzing vast amounts of data to optimize resources, manage traffic, enhance healthcare, and improve public safety. AI also supports energy management, environmental monitoring, and disaster response. Additionally, it facilitates smart governance and citizen engagement through efficient, transparent, and personalized services, making cities more liable and resilient. AI-driven systems are now pivotal in smart city infrastructures, providing real-time data analysis and decision-making capabilities that improve the efficiency and effectiveness of urban services [2].

Despite AI's many benefits to urban management, its rapid adoption in surveillance and security systems raises significant concerns [3]. The primary issues revolve around privacy, data protection, and ethical implementation. AI technologies can process vast amounts of data, often including sensitive personal information, which poses a risk of misuse or breaches. Moreover, deploying AI in surveillance can lead to invasive monitoring practices, undermining the privacy and freedoms of citizens. These concerns necessitate a comprehensive examination of the existing literature to understand the challenges and develop frameworks that balance technological benefits with ethical standards [4–6].

This research aims to conduct a bibliometric analysis to delve deeper into smart cities' security and surveillance dimensions, focusing on critical issues such as data protection and the ethical use of AI. By reviewing a comprehensive set of 745 articles indexed in Scopus from 1977 to 2023, the study employs co-citation analysis, co-occurrence analysis, and bibliographic coupling to explore the predominant themes and gaps in the existing literature. The goal is to map the research landscape and identify influential publications and key thematic clusters that shape the discourse around AI-driven smart city security and surveillance systems.

The study seeks to address the following research questions: What is the primary influence that characterizes variables (keywords) in the field of smart cities, and what are the interrelationships among them? What are the main thematic clusters that have evolved in research on smart cities, and how do these reflect the progress of the field? Which publications and associations have the most significant influence on smart city assessment among the extant literature? Corresponding to these questions, the objectives are to analyze the major factors that have the most significant impact on keywords in the smart city field, to discover correlations, characteristics, and relationship patterns among terms using cluster analysis, and to identify the articles and associations that have the most impact on the assessment of smart cities.

A notable gap exists in the current literature regarding the ethical implications and privacy concerns associated with deploying AI-powered security and surveillance technologies in smart cities. While numerous studies have examined the broader technology developments and the general concept of smart cities, there is a lack of focused research on AI-driven security frameworks and solutions. This omission highlights the urgent need for detailed academic inquiries that ensure advancements in AI do not compromise ethical standards and personal privacy [7–9].

The integration of AI in smart cities and its potential impact on governance, decision-making, and urban management has been a significant topic of discussion and exploration. Understanding the interplay between technological advancements and ethical considerations is paramount, given the critical role of security and surveillance systems in urban environments. This study aims to provide a thorough bibliometric analysis, combining qualitative and quantitative approaches to examine the scope of smart city security and surveillance [10, 11].

This research underscores the importance of a balanced approach to technological integration in smart cities, ensuring that advancements in AI and surveillance are aligned with ethical standards and community values. The findings are intended to guide urban planners and policymakers in developing strategies that enhance urban safety and efficiency, respect and protect individual privacy, and foster public trust. The study aims to contribute to the secure and sustainable growth of smart cities in the digital era by addressing the complexities and interconnections involved in deploying AI-powered technologies.

2 Methodology

Bibliometric analysis assesses the development of a specific topic of study by utilizing statistical methodologies [12]. According to Broadus [13], it assesses the amount, quality, importance, and impact of research outputs. Furthermore, the approach not only maps the contributions of researchers but also decrypts their citation networks. Moreover, it produces comprehensive visual representations of scientific findings and the theoretical foundations of the field. This analytical tool is essential for aiding the production of academic knowledge, tracking progress within a study topic, and identifying areas that require more inquiry [12]. To address the queries posited by our study, we will adhere to the procedural schematic presented in the forthcoming flow chart, as seen in Fig. 1.

2.1 Data Selection

The preliminary stage begins with identifying the data sources, with the Scopus database being used as the primary dataset. The process involves creating search terms that are applied to narrow down and sort the obtained data. This analysis used a meticulous and intentional choice of terms to create the search phrases. These phrases include "smart city," "security," "surveillance," and especially "AI." Together, these terms form the basis of the search criteria, aiming to extract a wide range of information from abstracts and whole publications. As a result of the method used, a total of 745 variables obtained solely from Scopus were discovered. This data collection improves the general understanding of the intersection between technological advancement and urban management.

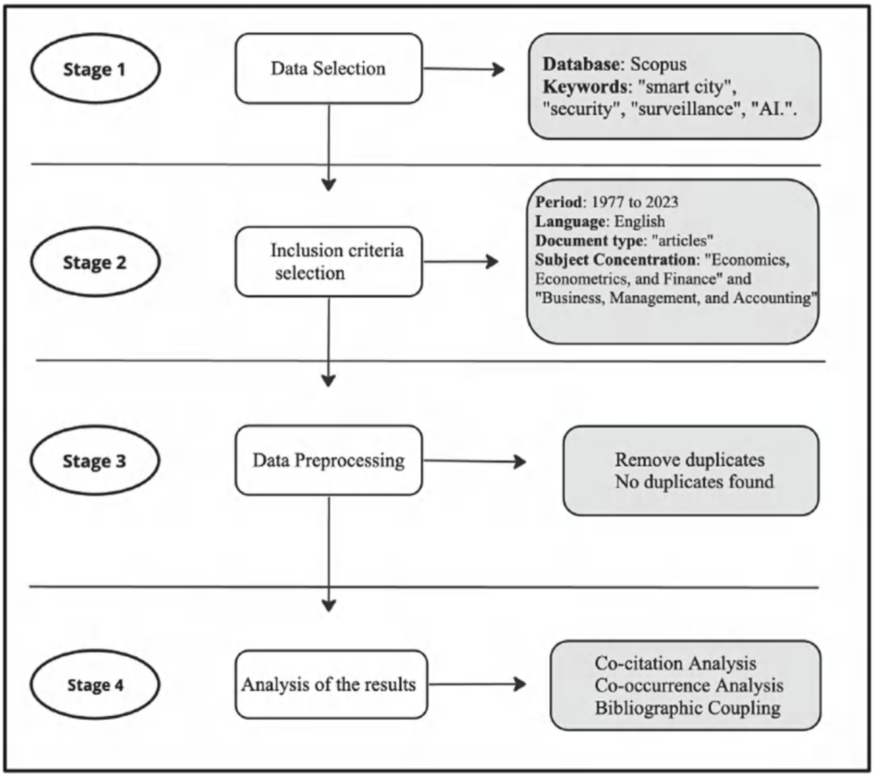


Fig. 1 Bibliometric methodology stages

The obtained data is the result of a comprehensive filtering strategy. At the outset, the subject concentration was narrowed to only the "Economics, Econometrics, and Finance" and "Business, Management, and Accounting" domains. This specific enhancement guarantees pertinence to smart city advancement's managerial and economic dimensions. These fields are integral to understanding the financial, managerial, and economic aspects of smart city advancements. Addressing this domain in the context of smart cities provides insights into the economic impacts, funding mechanisms, and cost-benefit analyses essential for smart city projects. Meanwhile, this domain offers perspectives on the management strategies, business models, and organizational frameworks that drive the implementation and sustainability of smart city initiatives. By concentrating on these areas, the study ensures a comprehensive understanding of how economic principles and business practices contribute to the successful development and operation of smart cities.

The subsequent procedures in our research approach involved the exclusive selection of 'articles' as the document type to ensure a rigorous academic standard. Finally, we restricted our inquiry exclusively to published final articles in academic journals, thereby guaranteeing that our evaluation is founded upon explicitly peer-reviewed and published scholarly works. By employing these criteria, the most relevant and credible sources were extracted, thereby facilitating an exhaustive examination of how AI can augment security and surveillance in the context of smart cities.

2.2 Data Preprocessing

Before executing our comprehensive bibliometric studies, a vital step in our methodology entailed rigorous data preprocessing to confirm the integrity and precision of our dataset. This involved thoroughly investigating replicated data entries, which could distort the findings and result in erroneous conclusions. Through our in-depth examination of the dataset, we found no duplication cases, proving the distinctiveness of each data item. By ensuring the absence of duplicates, we ensured the reliability of our analyses [14, 15].

2.3 Analysis Process

Our methodology incorporates an integrated bibliometric approach using Vosviewer to comprehensively examine the smart city research domain, blending qualitative depth with quantitative robustness. VOSviewer aids in the analysis and visualization of bibliometric data, making it easier to interpret complex relationships and trends within scientific literature. In this study we used VOSviewer to create and visualize bibliometric networks. These networks include journals, researchers, or individual publications which are constructed based on co-citation, or co-authorship relationships. We begin our analysis with a co-citation analysis to identify the most prolific authors and determine the most frequently cited journals within the field. This highlights key contributors and influential publications shaping the current landscape of smart city research. Our methodology extends to bibliographic coupling analysis, which allows us to explore the interconnectedness of studies through shared references, thereby highlighting the main thematic areas and contributing to a deeper understanding of the field's structure. Through this analysis, we categorize documents that pertain to separate yet interrelated research domains, providing a network perspective on the intellectual dialogue within the field.

In addition, bibliographic coupling and co-citation analysis were implemented to enhance our comprehension of the scholarly interface within smart city research. By quantifying the number of references shared between pairings of documents, bibliographic coupling determines their intellectual proximity. In contrast, co-citation analysis investigates the frequency at which a combination of documents are referenced concurrently, revealing the associative connections that connect distinct studies within the fundamental academic discourse.

The process is completed with a qualitative content analysis of the most cited papers, which adds interpretive depth to the quantitative results. This analysis thoroughly examined the scholarly contributions, delving into the intricacies and contextual foundations of the most influential work in the field. In addition, it facilitated the identification of research gaps, potential biases, and the general progression of the field's development.

3 Findings

Assembled collectively, these analytical elements constituted an all-encompassing bibliometric investigation, methodically delineating the intellectual domain of smart city research and illuminating its most impactful contributions and thematic trajectories. By evaluating the most referenced authors and works and identifying key research clusters through keyword analysis, this study aims to shed light on the current and future trends in AI and surveillance technologies. This exploration helps clarify the significant contributors and publications and the broader implications of these technologies on privacy, governance, and urban living. This comprehensive investigation is vital for navigating the complex challenges and possibilities that advanced technologies introduce into daily life.

Figure 2 illustrates a marked rise in the volume of scholarly publications over time. In recent years, we have witnessed a pronounced surge in focus on the convergence of smart city frameworks with the fields of business, management, accounting, economics, econometrics, and finance. This trend highlights the growing recognition of artificial intelligence's pivotal role in enhancing urban governance's security and surveillance facets. As a result, there is a notable shift toward more in-depth and specific scholarly research in this area, suggesting an increasing prioritization of AI's implications and applications in managing city dynamics and safety.

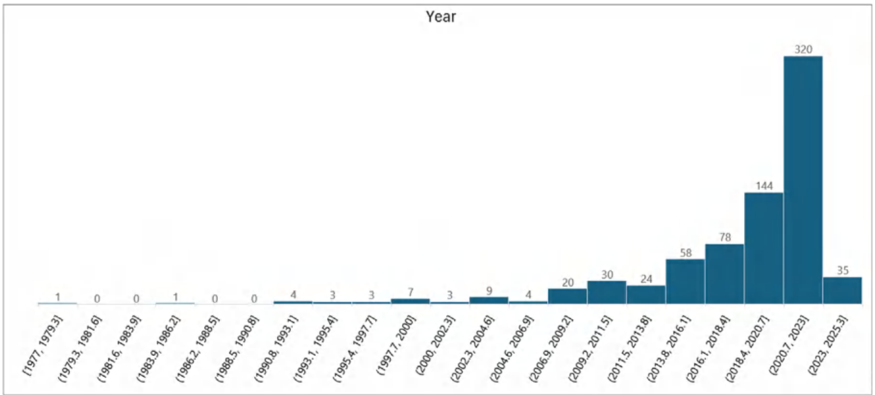


Fig. 2 Trend of bibliometric studies in AI-driven smart city security and surveillance

3.1 Descriptive Analysis

A descriptive analysis of the collected data provides an overview of the publication trends, showing a significant increase in the number of publications on AI and smart cities over the past two decades. This trend indicates growing academic interest and the increasing relevance of these topics. This section is devoted to presenting the data accordingly.

3.2 Co-citation Analysis

The co-citation analysis provided significant insights into the most influential contributors and seminal works within the AI-driven smart city security and surveillance systems field. By examining the frequency with which two documents are cited, the study highlighted key authors and journals that shape the discourse in this area. The analysis identified the most contributed authors based on their co-citation frequencies, revealing their centrality and influence in the research network. As detailed in Table 1, these authors have made substantial contributions, with their works frequently cited together, indicating their pivotal role in developing this field.

Additionally, the analysis of the top 10 most cited journals, presented in Table 2, showcased the primary sources of influential research. These journals, which include prominent publications in technology and urban studies, have been instrumental in disseminating key findings and theories related to AI applications in smart city environments. Their high citation counts underscore their importance as critical knowledge and innovation repositories.

The co-citation analysis, illustrated in Fig. 3, reveals that authors like Kirstie Ball and Douglas Cumming are pivotal, with their works being central to discussions on surveillance as a marketplace icon and global market surveillance, respectively. Furthermore, journals like the Journal of Information Technology have been instrumental in disseminating pioneering research by scholars such as Zuboff, whose work on surveillance capitalism is extensively cited.

Table 1 The most contributed author

Ranking	Author	Documents	Citations	Avg. citations
1	Ball, Kirstie	6	113	18.8333
2	Cumming, Douglas	5	112	22.4
3	Kajalo, Sami	5	72	14.4

Table 2 Top 10 most cited journals

Authors	Title	Year	Source title	Cited
Zuboff [16]	Big other: surveillance capitalism and the prospects of an information civilization	2015	Journal of Information Technology	2981
Allam and Dhunny [17]	On big data, artificial intelligence and smart cities	2019	Cities	988
Pandey and Pal [18]	Impact of digital surge during Covid-19 pandemic: a viewpoint on research and practice	2020	International Journal of Information Management	890
Dwivedi et al. [19]	So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy	2023	International Journal of Information Management	688
Kaur and Rampersad [20]	Trust in driverless cars: investigating key factors influencing the adoption of driverless cars	2018	Journal of Engineering and Technology Management—JET-M	327
Greenberg and Barling [21]	Predicting employee aggression against coworkers, subordinates and supervisors: the roles of person behaviors and perceived workplace factors	1999	Journal of Organizational Behavior	220
Rao et al. [22]	The societal impact of commercial drones	2016	Technology in Society	213
Liu et al. [23]	Deep convolutional neural networks for thermal infrared object tracking	2017	Knowledge-Based Systems	174
Jain et al. [24]	Software assistants for randomized patrol planning for the lax airport police and the Federal Air Marshal Service	2010	Interfaces	153
Gennotte and Pyle [25]	Capital controls and bank risk	1991	Journal of Banking and Finance	152

3.3 Co-occurrence Analysis

The co-occurrence analysis focused on identifying the main themes and topics within the literature by examining the frequency and patterns of keyword appearances. Using general and author-provided keywords, this analysis provided a comprehensive overview of the prevalent research areas and their interconnections [26]. The co-occurrence analysis using general keywords, visualized in Fig. 4, and shown in Table 3, mapped the relationships between frequently used terms such as "AI," "smart

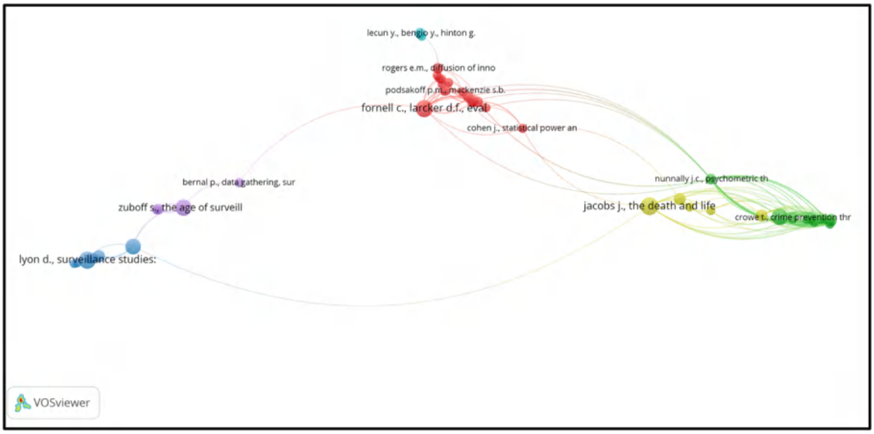


Fig. 3 Co-citation analysis graph

city," "security," and "privacy." This network map illustrated the central themes and their interrelations, indicating the multifaceted nature of the research in this domain.

In the second part of our analysis, we employ keyword co-occurrence techniques to map the central themes and trends within the research community. Our analysis, displayed in Fig. 4, identifies distinct clusters such as "Security," "Surveillance," and "Privacy," each representing a comprehensive synthesis of related concepts and challenges in smart city security and surveillance. This approach categorizes the literature into groups that share common themes and delineates the boundaries of emergent and declining research trends.

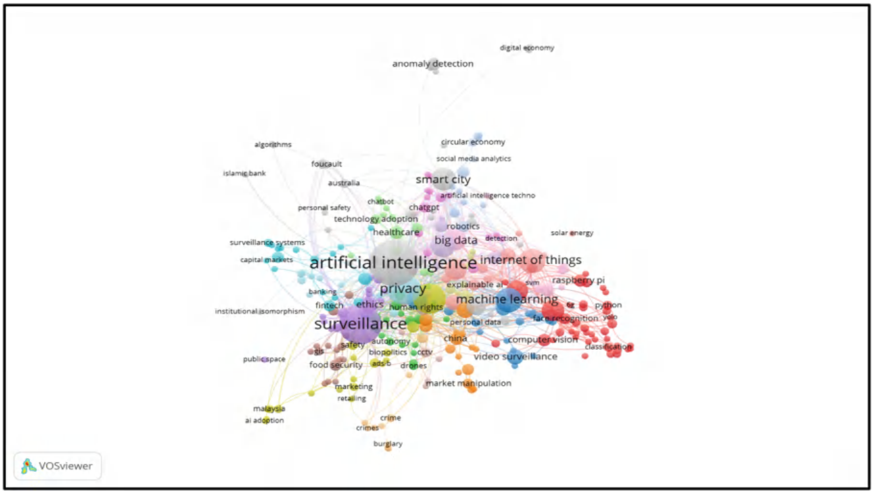


Fig. 4 Co-occurrence analysis using keywords

Table 3 Co-occurrence cluster with keywords

Cluster	Title (occurrence)	Keywords
1	Deep learning (21)	Deep learning, IoT, Raspberry Pi, computer vision, face recognition, fog computing, convolutional neural network, python, classification, surveillance system, OpenCV, vehicle detection, CNN, face detection, object detection, wireless sensor networks, motion detection, support vector machine, outlier detection, surveillance cameras, image processing, cameras, 6G, YOLO, environmental monitoring, IDS, security surveillance, cloud, botnet, object tracking, particle filter, smartphone, citizen science, feature selection, human tracking, MQTT, PID controller, visual surveillance, background subtraction, Gaussian mixture model, smartphones
2	Resilience (6)	Resilience, law, European Union, governance, algorithm, autonomy, UAVS, drones, bias, discrimination, intelligence, ethic, policy, prioritization, privacy by design, public policy, covid, robot, vaccine
3	Machine learning (29)	Machine learning, covid-19, video surveillance, smart cities, mass surveillance, framework, personal data, biometric, authentication, network security, digital surveillance, acceptance, data management, energy optimization, WSN, transfer learning, unmanned, aerial vehicle, control
4	Security (42)	Security, surveillance capitalism, safety, data mining, marketing, PLS-SEM, Malaysia, biopolitics, retail, retailing, consumer behaviour, customer experience, qualitative research, ADS-B, GPS, talent acquisition, border security, AI adoption
5	Surveillance (71)	Surveillance, ethics, terrorism, power, metadata, identity, gender, interception, data retention, deterrence, secrecy, whistleblowing, corporate security, data retention directive, public security, travel, ontology, public space
6	Trust (12)	Trust, information technology, fintech, Delphi, financial services, surveillance systems, structural equation modeling, human–computer interaction, mobile marketing, services marketing, banking, service quality, banking industry, challenges, technology acceptance model, social sorting, regulations, capital markets

(continued)

Table 3 (continued)

Cluster	Title (occurrence)	Keywords
7	Cybersecurity (12)	Cybersecurity, technology, regulation, China, insider trading, market manipulation, risk management, securities regulation, Russia, economic development, law and finance, airport security, enforcement, EU, modelling, securities
8	Food security (6)	Food security, transparency, GIS, decision making, renewable energy, uncertainty, social control, technologies, camera, complexity, network analysis, remote sensing, agriculture, information and communication technology, management, values
9	ChatGPT (5)	ChatGPT, India, disruptive technology, fundamental rights, generative AI, food supply chain, supply chain management, data science, detection, generative artificial intelligence, organizational learning, emerging technologies, aeronautics, performance
10	Internet of things (29)	Internet of things, AI, internet of things (IoT), cloud computing, explainable AI, intrusion detection, GDPR, IoT security, big data analysis, neural networks, wireless sensor network, throughput, solar energy
11	Healthcare (8)	Healthcare, innovation, technology adoption, learning, cyber threat intelligence, chatbot, anthropomorphism, disinformation, education, finance, natural language processing, openness
12	Robotics (6)	Robotics, industry 4.0, digital transformation, circular economy, waste management, automation, chatbots, social media analytics, technostress, data analytics, social network analysis, sentiment analysis
13	Human rights (5)	Human rights, national security, information security, data, digital trade, intellectual property, securitization, human security, securitization, technological innovation, facial recognition
14	Artificial intelligence (AI) (26)	Artificial intelligence (AI), sustainability, cyber security, mobility, digitalization, machine learning (ML), confidentiality, cyberattacks, SVM, maritime security
15	Privacy (45)	Privacy, fraud, social media, internet, policing, the Netherlands, youth, Facebook, marketplace icon, security measures
16	Crime prevention (5)	Crime prevention, crime, customers, hotel, crimes, shoplifting, burglary, CPTED, situational crime prevention, theft

(continued)

Table 3 (continued)

Cluster	Title (occurrence)	Keywords
17	Data security (8)	Data security, CCTV, accountability, legitimacy, standards, sanctions, public sector, proportionality, closed circuit television, human factors
18	Big data (22)	Big data, biometrics, digital identity, automated identification, identity crime, identity surveillance, mobile identity, accuracy, artificial intelligence technology
19	Smart city (25)	Smart city, anomaly detection, smart surveillance, future frame prediction, intelligent video surveillance, responsible innovation, firefly algorithm, vehicle trajectory
20	Australia (4)	Australia, Foucault, disciplinary power, indirect government, normalization, fear of crime, built environment, personal safety
21	Smart home (4)	Smart home, health, privacy concerns, technology trust, technology acceptance, well-being, convolutional neural networks, internet-of-things (IoT)
22	Artificial intelligence (98)	Algorithms, mobile banking, communication, institutional isomorphism, Islamic bank
23	Data protection (15)	Data protection, public attitudes, automated decision-making, privacy perceptions
24	Digital economy (2)	Digital economy, Nigeria, strategy
25	Blockchain (15)	Blockchain, federated learning, AI ethics

The author keyword analysis, depicted in Fig. 5, further refined these insights by highlighting more specific areas of focus, such as “big data,” “urban surveillance,” and “ethical AI.” These visualizations revealed the detailed landscape of research interests and the interconnectedness of various topics.

3.4 Bibliographic Coupling

The bibliographic coupling analysis grouped documents based on shared references, forming clusters representing thematic areas within the research field, as shown in Table 4. This method provided a deeper understanding of the intellectual structure and thematic evolution of AI-driven smart city security and surveillance research.

Cluster 2, titled “The Multi-faceted Impact of AI Adoption,” examined the diverse impacts of AI integration in smart city initiatives. This cluster highlighted the benefits and challenges associated with AI technologies, emphasizing the need for a balanced approach to innovation and ethical considerations. Cluster 11, “Strategic Integration

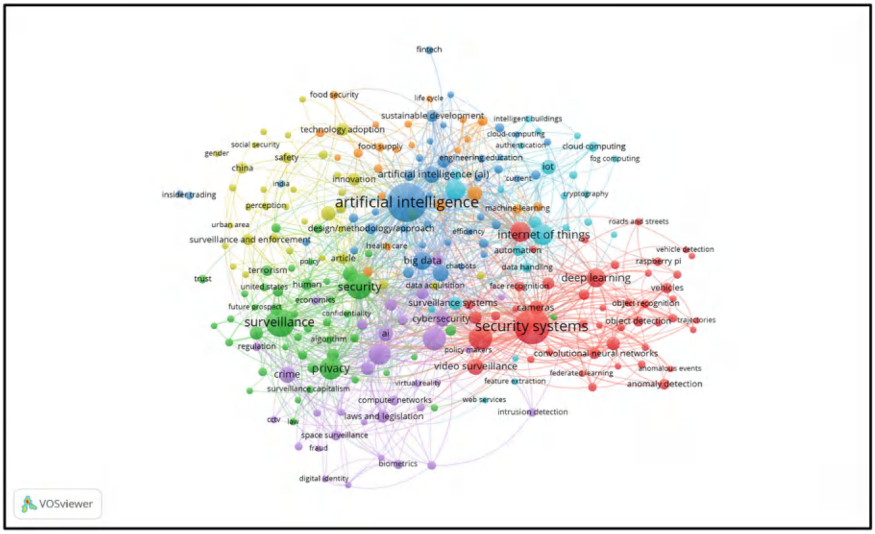


Fig. 5 Co-occurrence cluster using author keywords

of AI and Big Data in Smart City Development," focused on using AI and big data to enhance urban management and infrastructure. The studies in this cluster demonstrated how data analytics can drive improvements in city operations and service delivery.

Cluster 14, "The Socio-technical Dynamics of Emerging Technologies," explored the interaction between society and technology. This cluster underscored the importance of understanding the socio-technical dynamics that emerge from implementing new technologies in urban environments. Finally, Cluster 16, "The Technological Impacts on Society," delved into the broader societal impacts of technological advancements. This cluster mainly focused on surveillance, privacy, and urban dynamics, emphasizing the need to balance technological benefits with ethical and privacy concerns.

4 Discussion

The findings of this study provide a multifaceted understanding of the current landscape and evolving trends in AI-driven smart city security and surveillance systems. The research highlights significant contributions, central themes, and interconnected topics within the field by employing co-citation, co-occurrence, and bibliographic coupling analyses. These findings not only elucidate the academic discourse but also underscore critical considerations for the practical implementation of AI technologies in urban settings.

Table 4 Bibliographic Coupling Thematic Areas

Authors (year)	Article	Source	Citations
Cluster 1: The Critical Intersection of Surveillance, Security, and Society (46 articles)			
Dunn Cavelty [27]	Breaking the Cyber-Security Dilemma: Aligning Security Needs and Removing Vulnerabilities	Science and Engineering Ethics	72
Coudert et al. [28]	Body-worn cameras for police accountability: Opportunities and risks	Computer Law and Security Review	71
Wood, Ball [29]	Brandscapes of control? Surveillance, marketing and the co-construction of subjectivity and space in neo-liberal capitalism	Marketing Theory	69
Cluster 2: The Multi-Dimensional Impact of AI Adoption (44 articles)			
Dwivedi et al. [19]	"So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy	International Journal of Information Management	488
Greenberg, Barling [21]	Predicting employee aggression against coworkers, subordinates and supervisors: The roles of person behaviors and perceived workplace factors	Journal of Organizational Behavior	220
Pillai, Sivathanu [30]	Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations	Benchmarking	110
Cluster 3: Innovations in AI-Driven Surveillance (36 articles)			
Liu et al. [23]	Deep convolutional neural networks for thermal infrared object tracking	Knowledge-Based Systems	174
Pérez-Hernández et al. [31]	Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance	Knowledge-Based Systems	118
Chen et al. [32]	Video-Based Detection Infrastructure Enhancement for Automated Ship Recognition and Behavior Analysis	Journal of Advanced Transportation	96
Cluster 4: Surveillance, Safety, and Environmental Design (34 articles)			
Paydar et al. [33]	Perceived security of women in relation to their path choice toward sustainable neighborhood in Santiago, Chile	Cities	55

(continued)

Table 4 (continued)

Authors (year)	Article	Source	Citations
Welsh et al. [34]	Reconceptualizing public area surveillance and crime prevention: Security guards, place managers and defensible space	Security Journal	40
Priks [35]	The Effects of Surveillance Cameras on Crime: Evidence from the Stockholm Subway	Economic Journal	38
Cluster 5: Advancements in AI-Enabled Surveillance and Decision Support for Safety and Efficiency (33 articles)			
Yadav et al. [36]	A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions	Knowledge-Based Systems	111
Bhatt, Zaveri [37]	The enabling role of decision support systems in organizational learning	Decision Support Systems	73
Chory et al. [38]	Organizational Surveillance of Computer-Mediated Workplace Communication: Employee Privacy Concerns and Responses	Employee Responsibilities and Rights Journal	36
Cluster 6: IoT and Big Data's Influence on Privacy, Governance, and Performance in Smart Cities (31 articles)			
Caron et al. [39	The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective	Computer Law and Security Review	89
Javaid et al. [40]	Significant applications of big data in industry 4.0	Journal of Industrial Integration and Management	43
Hsiao et al. [41]	Elevated performance of the smart city-a case study of the IoT by innovation mode	IEEE Transactions on Engineering Management	36
Cluster 7: The Intersection of Artificial Intelligence with Cultural, Ethical, and Environmental Governance (29 articles)			
Robinson [42]	Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI)	Technology in Society	71
Winter, Davidson [43]	Governance of artificial intelligence and personal health information	Digital Policy, Regulation and Governance	40
Jupe, Keatley [44]	Airport artificial intelligence can detect deception: or am i lying?	Security Journal	18

(continued)

Table 4 (continued)

Authors (year)	Article	Source	Citations
Cluster 8: Navigating the Complex Interface of Technology, Surveillance, and Governance in Contemporary Society. (27 articles)			
Rowe [45]	Contact tracing apps and values dilemmas: A privacy paradox in a neo-liberal world	International Journal of Information Management	117
Ribeiro-Navarrete et al. [46]	Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy	Technological Forecasting and Social Change	115
Spanaki et al. [47]	Artificial intelligence and food security: swarm intelligence of AgriTech drones for smart AgriFood operations	Production Planning and Control	36
Cluster 9: Integrating AI across Domains: Enhancing Trust, Transparency, and Innovation (27 articles)			
Nilashi et al. [48]	The role of Security, Design and Content factors on customer trust in mobile commerce	Journal of Retailing and Consumer Services	100
Yigitcanlar et al. [49]	Responsible urban innovation with local government artificial intelligence (Ai): A conceptual framework and research agenda	Journal of Open Innovation: Technology, Market, and Complexity	84
Raza et al. [50]	Designing ECG monitoring healthcare system with federated transfer learning and explainable AI	Knowledge-Based Systems	69
Cluster 10: AI in Business: Enhancing Resilience and Innovation Across Sectors 27 (articles)			
Cubric [51]	Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study	Technology in Society	122
Königstorfer, Thalmann [52]	Applications of Artificial Intelligence in commercial banks—A research agenda for behavioral finance	Journal of Behavioral and Experimental Finance	81
Modgil et al. [53]	AI technologies and their impact on supply chain resilience during—19	International Journal of Physical Distribution and Logistics Management	79
Cluster 11: Strategic Integration of AI and Big Data in Smart City Development (25 articles)			
Allam, Dhunny [17]	On big data, artificial intelligence and smart cities	Cities	562
Yigitcanlar et al. [54]	Artificial intelligence technologies and related urban planning and development concepts: How are they perceived and utilized in Australia?	Journal of Open Innovation: Technology, Market, and Complexity	86

(continued)

Table 4 (continued)

Authors (year)	Article	Source	Citations
Herath, Mittal [3]	Adoption of artificial intelligence in smart cities: A comprehensive review	International Journal of Information Management Data Insights	64
Cluster 12: Consumer Adoption and Trust in AI-Enabled Technologies (21 articles)			
Kaur, Rampersad [55]	Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars	Journal of Engineering and Technology Management - JET-M	327
Kowalczuk [56]	Consumer acceptance of smart speakers: a mixed methods approach	Journal of Research in Interactive Marketing	110
Gansser, Reich [57]	A new acceptance model for artificial intelligence with extensions to UTAUT2: An empirical study in three segments of application	Technology in Society	92
Cluster 13: Navigating the Societal and Regulatory Landscapes of Emerging Technologies (21 articles)			
Rao et al. [22]	The societal impact of commercial drones	Technology in Society	213
Weber [58]	Internet of things: Privacy issues revisited	Computer Law and Security Review	141
Ding et al. [59]	Smart logistics based on the internet of things technology: an overview	International Journal of Logistics Research and Applications	138
Cluster 14: The Socio-Technical Dynamics of Emerging Technologies (19 articles)			
Zuboff [16]	Big other: Surveillance capitalism and the prospects of an information civilization	Journal of Information Technology	1841
Broeders et al. [60]	Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data	Computer Law and Security Review	54
Montgomery [61]	Youth and surveillance in the Facebook era: Policy interventions and social implications	Telecommunications Policy	45
Cluster 15: Enhancing Financial and Healthcare Surveillance through Advanced Technologies (16 articles)			
Gómez-Vallejo et al. [62]	A case-based reasoning system for aiding detection and classification of nosocomial infections	Decision Support Systems	53
Cumming et al. [63]	Same rules, different enforcement: Market abuse in Europe	Journal of International Financial Markets, Institutions and Money	34
Chen (2020b)	IQ from IP: Simplifying search in portfolio choice	Journal of Financial Economics	32

(continued)

Table 4 (continued)

Authors (year)	Article	Source	Citations
Cluster 16: Technological Impacts on Society: Surveillance, Privacy, and Urban Dynamics (16 articles)			
Pandey, Pal [18]	Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice	International Journal of Information Management	530
Cybriwsky [64]	Changing patterns of urban public space. Observations and assessments from the Tokyo and New York metropolitan areas	Cities	106
Anteby, Chan [65]	A self-fulfilling cycle of coercive surveillance: Workers' invisibility practices and managerial justification	Organization Science	76
Cluster 17: Navigating Technological and Regulatory Transformations in Diverse Sectors (10 articles)			
Nienaber et al. [66]	Do we bank on regulation or reputation? A meta-analysis and meta-regression of organizational trust in the financial services sector	International Journal of Bank Marketing	42
De Keyser et al. [67]	Opportunities and challenges of using biometrics for business: Developing a research agenda	Journal of Business Research	20
Li et al. [68]	Mapping two decades of smart home research: A systematic scientometric analysis	Technological Forecasting and Social Change	19
Cluster 18: Enhancing Safety and Security in Digital and Urban Environments (2)			
Elhai et al. [69]	Anxiety about electronic data hacking: Predictors and relations with digital privacy protection behavior	Internet Research	25
Li et al. [70]	CPT Model-Based Prediction of the Temporal and Spatial Distributions of Passenger Flow for Urban Rail Transit under Emergency Conditions	Journal of Advanced Transportation	7

4.1 Key Contributors and Influential Journals

The co-citation analysis reveals that specific authors and journals hold significant sway in shaping AI-driven smart city security research. Notably, authors such as Shoshana Zuboff and Allam Z. have been frequently cited together, indicating their seminal contributions to understanding AI's ethical, social, and technical dimensions in urban environments. The prominence of journals like "Journal of Information Technology" and "Sustainability" underscores their role as critical platforms for disseminating influential research. These sources serve as vital knowledge repositories, offering a comprehensive overview of the challenges and opportunities AI presents in smart cities.

4.2 Central Themes and Research Trends

The co-occurrence analysis identifies key themes dominating the smart city security discourse. Terms like "AI," "smart city," "security," and "privacy" are central, reflecting the core concerns of researchers. The frequent appearance of "big data" and "urban surveillance" in author keywords further highlights the importance of data management and ethical surveillance practices. These findings indicate a research focus on leveraging AI to enhance urban security while addressing privacy concerns. The intricate balance between utilizing advanced technologies for public safety and safeguarding individual rights emerges as a critical theme.

4.3 Thematic Clusters and Intellectual Structure

The bibliographic coupling analysis groups related studies into distinct clusters, each representing a specific aspect of AI-driven smart city research. Cluster 2, "The Multi-faceted Impact of AI Adoption," explores the diverse effects of AI technologies, from operational efficiencies to ethical dilemmas. This cluster underscores the dual-edged nature of AI, capable of both significant advancements and potential ethical pitfalls.

Cluster 11, "Strategic Integration of AI and Big Data in Smart City Development," highlights the transformative potential of data analytics in urban management. Studies within this cluster demonstrate how AI and big data can optimize city operations, enhance service delivery, and drive sustainable development. However, they also stress the need for strategic planning and ethical considerations to prevent misuse and ensure equitable benefits. Cluster 14, "The Socio-technical Dynamics of Emerging Technologies," delves into the complex interactions between society and technology. This cluster emphasizes that technological implementations in smart cities are not merely technical challenges but also social ones. Understanding these dynamics is crucial for ensuring technological advancements align with societal values and norms.

Cluster 16, "The Technological Impacts on Society," focuses on the broader societal implications of AI and surveillance technologies. This cluster addresses concerns about privacy, data security, and the ethical deployment of surveillance systems. The studies highlight the necessity of transparent policies, robust data protection measures, and continuous ethical oversight to maintain public trust and protect individual rights.

5 Contribution

This study makes substantial contributions to the field of AI-driven smart city security by providing a comprehensive mapping of the research landscape, identifying key themes and trends, addressing ethical and privacy concerns, offering policy recommendations, bridging academic research with practical application, laying a foundation for future research, and emphasizing the importance of public trust and engagement. These contributions collectively advance the understanding and implementation of AI technologies in creating secure, efficient, and ethically sound smart cities.

The primary contributions include providing a detailed mapping of the research landscape through co-citation, co-occurrence, and bibliographic coupling analyses. This approach identifies central themes, key authors, and influential journals, offering a holistic field view. This mapping is a valuable resource for researchers seeking to understand the current state of AI-driven smart city security and identify potential areas for future investigation. The co-occurrence analysis highlights critical themes such as AI, big data, urban surveillance, and privacy, reflecting contemporary research's core concerns and focus areas. The study elucidates the primary issues researchers and practitioners must address by identifying these recurring themes. This thematic analysis provides a foundation for developing targeted research agendas and practical solutions. The study underscores the ethical and privacy concerns associated with AI-driven surveillance technologies. By examining clusters focusing on socio-technical dynamics and technological impacts on society, the research highlights the necessity of balancing technological advancements with ethical considerations. This insight is crucial for policymakers and urban planners tasked with implementing AI technologies in a manner that respects individual rights and societal values.

Drawing from the analysis of thematic clusters, the study offers policy and governance recommendations to ensure the ethical deployment of AI technologies in smart cities. These recommendations emphasize the importance of comprehensive legal frameworks, transparency in data usage, public engagement, and continuous ethical oversight. These contributions are vital for developing policies that foster public trust and protect individual privacy while leveraging AI for urban security. The study bridges the gap between academic research and practical application by providing actionable insights for urban planners, policymakers, and technology developers. By highlighting the practical implications of research findings, the study ensures that theoretical advancements are translated into real-world solutions. This approach enhances the relevance and impact of academic research in smart city security.

The bibliometric analysis conducted in this study lays a solid foundation for future research. The study guides researchers toward unexplored areas and potential research gaps by identifying influential works, key themes, and emerging trends. This contribution is essential for advancing the field and fostering innovative approaches to AI-driven smart city security and surveillance. Finally, the study emphasizes the importance of public trust and engagement in deploying AI technologies. The research promotes a participatory approach to smart city development by advocating for transparent communication and public consultations. This contribution is critical for ensuring that technological advancements align with public expectations and contribute positively to urban life.

6 Conclusion

This study has conducted a comprehensive bibliometric analysis to explore the landscape of AI-driven smart city security and surveillance systems. The research identified critical themes, influential authors, and key publications that shape the field by mapping the co-citation, co-occurrence, and bibliographic coupling clusters. The findings highlight the central concerns of privacy, ethical deployment, and public trust, emphasizing the need for robust legal frameworks and transparent data practices.

The study underscores the necessity of enhancing urban security through advanced AI technologies while safeguarding individual rights and societal values. Policy recommendations are crucial to ensure that AI implementations in smart cities are practical and ethically sound. Additionally, the research bridges the gap between theoretical advancements and practical applications, offering actionable insights for policymakers, urban planners, and technology developers. The study contributes significantly to the field by offering a detailed overview of the current research landscape, addressing key ethical and privacy issues, and providing practical recommendations for deploying AI technologies in smart cities. These contributions are essential for fostering the development of secure, efficient, and ethically responsible urban environments, paving the way for future research and innovation.

However, one limitation of our study is the potential biases inherent in the Scopus database. Scopus, while comprehensive, may not include all relevant publications, particularly those from non-English language sources or smaller regional journals, which could skew the results. Additionally, our focus on the "Economics, Econometrics, and Finance" and "Business, Management, and Accounting" domains may exclude significant research from other relevant fields such as engineering or computer science. The exclusion of certain document types, such as conference papers or grey literature, could also lead to an incomplete representation of the literature. These limitations may affect the findings by potentially overlooking diverse perspectives and recent advancements in smart city research from other disciplines or publication types.

References

1. Zhang X, Manogaran G, Muthu B. IoT enabled integrated system for green energy into smart cities. Sustainable Energy Technologies and Assessments. 2021;46:101208.
2. Ye X, Newman G, Lee C, Van Zandt S, Jourdan D. Toward Urban artificial intelligence for developing justice-oriented smart cities. SAGE Publications Sage CA: Los Angeles, CA; 2023. p. 6–7.

3. Herath H, Mittal M. Adoption of artificial intelligence in smart cities: A comprehensive review. International Journal of Information Management Data Insights. 2022;2(1):100076.
4. Shahriar S, Allana S, Hazratifard SM, Dara R. A survey of privacy risks and mitigation strategies in the Artificial Intelligence life cycle. IEEE Access. 2023;11:61829–54.
5. Lannon C, Nelson J, Cunneen M. Ethical AI for automated bus lane enforcement. Sustainability. 2021;13(21):11579.
6. Almeida D, Shmarko K, Lomas E. The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. AI and Ethics. 2022;2(3):377–87.
7. Szpilko D, Naharro FJ, Lăzăroiu G, Nica E, de la Torre Gallegos A. Artificial intelligence in the smart city—a literature review. Engineering Management in Production and Services. 2023;15(4):53–75.
8. Ismagilova E, Hughes L, Rana NP, Dwivedi YK. Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework. Information Systems Frontiers. 2022:1–22.
9. Ardabili BR, Pazho AD, Noghre GA, Neff C, Bhaskararayuni SD, Ravindran A, et al. Understanding policy and technical aspects of ai-enabled smart video surveillance to address public safety. Computational Urban Science. 2023;3(1):21.
10. Alaeddini M, Hajizadeh M, Reaidy P. A Bibliometric Analysis of Research on the Convergence of Artificial Intelligence and Blockchain in Smart Cities. Smart Cities. 2023;6(2):764–95.
11. Yigitcanlar T, Cugurullo F. The sustainability of artificial intelligence: An urbanistic viewpoint from the lens of smart and sustainable cities. Sustainability. 2020;12(20):8548.
12. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. Journal of business research. 2021;133:285–96.
13. Broadus RN. Toward a definition of "bibliometrics". Scientometrics. 1987;12:373–9.
14. Tarun S, Batth RS, Kaur S. A scheme for data deduplication using advance machine learning architecture in distributed systems. 2021 International Conference on Computing Sciences (ICCS): IEEE; 2021. p. 53–60.
15. Sandu A, Ioanăș I, Delcea C, Florescu M-S, Cotfas L-A. Numbers Do Not Lie: A Bibliometric Examination of Machine Learning Techniques in Fake News Research. Algorithms. 2024;17(2):70.
16. Zuboff S. Big other: surveillance capitalism and the prospects of an information civilization. Journal of information technology. 2015;30(1):75–89.
17. Allam Z, Dhunny ZA. On big data, artificial intelligence and smart cities. Cities. 2019;89:80–91.
18. Pandey N, Pal A. Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. International journal of information management. 2020;55:102171.
19. Dwivedi YK, Kshetri N, Hughes L, Slade EL, Jeyaraj A, Kar AK, et al. Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management. 2023;71:102642.
20. Kaur K, Rampersad G. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. Journal of Engineering and Technology Management. 2018;48:87–96.
21. Greenberg L, Barling J. Predicting employee aggression against coworkers, subordinates and supervisors: The roles of person behaviors and perceived workplace factors. Journal of Organizational Behavior. 1999;20(6):897–913.
22. Rao B, Gopi AG, Maione R. The societal impact of commercial drones. Technology in society. 2016;45:83–90.
23. Liu Q, Lu X, He Z, Zhang C, Chen W-S. Deep convolutional neural networks for thermal infrared object tracking. Knowledge-Based Systems. 2017;134:189–98.
24. Jain M, Tsai J, Pita J, Kiekintveld C, Rathi S, Tambe M, et al. Software assistants for randomized patrol planning for the lax airport police and the federal air marshal service. Interfaces. 2010;40(4):267–90.
25. Gennotte G, Pyle D. Capital controls and bank risk. Journal of Banking & Finance. 1991;15(4–5):805–24.
26. Zhao L, Tang Z-y, Zou X. Mapping the knowledge domain of smart-city research: A bibliometric and scientometric analysis. Sustainability. 2019;11(23):6648.

27. Dunn Cavelty M. Breaking the cyber-security dilemma: Aligning security needs and removing vulnerabilities. Science engineering ethics. 2014;20:701–15.
28. Coudert F, Butin D, Le Métayer D. Body-worn cameras for police accountability: Opportunities and risks. Computer law security review. 2015;31(6):749–62.
29. Wood DM, Ball K. Brandscapes of control? Surveillance, marketing and the co-construction of subjectivity and space in neo-liberal capitalism. Marketing Theory. 2013;13(1):47–67.
30. Pillai R, Sivathanu B. Adoption of artificial intelligence (AI) for talent acquisition in IT/ITeS organizations. Benchmarking: An International Journal. 2020;27(9):2599–629.
31. Pérez-Hernández F, Tabik S, Lamas A, Olmos R, Fujita H, Herrera F. Object detection binary classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance. Knowledge-Based Systems. 2020;194:105590.
32. Chen X, Qi L, Yang Y, Luo Q, Postolache O, Tang J, et al. Video-Based Detection Infrastructure Enhancement for Automated Ship Recognition and Behavior Analysis. Journal of Advanced Transportation. 2020;2020(1):7194342.
33. Paydar M, Kamani-Fard A, Etminani-Ghasrodashti R. Perceived security of women in relation to their path choice toward sustainable neighborhood in Santiago, Chile. Cities. 2017;60:289–300.
34. Welsh BC, Mudge ME, Farrington DP. Reconceptualizing public area surveillance and crime prevention: Security guards, place managers and defensible space. Security Journal. 2010;23:299–319.
35. Priks M. The effects of surveillance cameras on crime: Evidence from the Stockholm subway. The Economic Journal. 2015;125(588):F289–F305.
36. Yadav SK, Tiwari K, Pandey HM, Akbar SA. A review of multimodal human activity recognition with special emphasis on classification, applications, challenges and future directions. Knowledge-Based Systems. 2021;223:106970.
37. Bhatt GD, Zaveri J. The enabling role of decision support systems in organizational learning. Decision Support Systems. 2002;32(3):297–309.
38. Chory RM, Vela LE, Avtgis TA. Organizational surveillance of computer-mediated workplace communication: Employee privacy concerns and responses. Employee Responsibilities Rights Journal. 2016;28:23–43.
39. Caron X, Bosua R, Maynard SB, Ahmad A. The Internet of Things (IoT) and its impact on individual privacy: An Australian perspective. Computer law security review. 2016;32(1):4–15.
40. Javaid M, Haleem A, Singh RP, Suman R. Significant applications of big data in Industry 4.0. Journal of Industrial Integration Management. 2021;6(04):429–47.
41. Hsiao Y-C, Wu M-H, Li SC. Elevated performance of the smart city—a case study of the IoT by innovation mode. IEEE Transactions on Engineering Management. 2019;68(5):1461–75.
42. Robinson SC. Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). Technology in Society. 2020;63:101421.
43. Winter JS, Davidson E. Governance of artificial intelligence and personal health information. Digital policy, regulation governance. 2019;21(3):280–90.
44. Jupe LM, Keatley DA. Airport artificial intelligence can detect deception: or am i lying? Security Journal. 2020;33(4):622–35.
45. Rowe F. Contact tracing apps and values dilemmas: A privacy paradox in a neo-liberal world. International Journal of Information Management. 2020;55:102178.
46. Ribeiro-Navarrete S, Saura JR, Palacios-Marqués D. Towards a new era of mass data collection: Assessing pandemic surveillance technologies to preserve user privacy. Technological Forecasting Social Change. 2021;167:120681.
47. Spanaki K, Karafili E, Sivarajah U, Despoudi S, Irani Z. Artificial intelligence and food security: swarm intelligence of AgriTech drones for smart AgriFood operations. Production Planning Control. 2022;33(16):1498–516.
48. Nilashi M, Ibrahim O, Mirabi VR, Ebrahimi L, Zare M. The role of Security, Design and Content factors on customer trust in mobile commerce. Journal of Retailing Consumer Services. 2015;26:57–69.
49. Yigitcanlar T, Corchado JM, Mehmood R, Li RYM, Mossberger K, Desouza K. Responsible urban innovation with local government artificial intelligence (AI): A conceptual framework and research agenda. Journal of Open Innovation: Technology, Market, Complexity. 2021;7(1):71.

50. Raza A, Tran KP, Koehl L, Li S. Designing ECG monitoring healthcare system with federated transfer learning and explainable AI. Knowledge-Based Systems. 2022;236:107763.
51. Cubric M. Drivers, barriers and social considerations for AI adoption in business and management: A tertiary study. Technology in Society. 2020;62:101257.
52. Königstorfer F, Thalmann S. Applications of Artificial Intelligence in commercial banks–A research agenda for behavioral finance. Journal of behavioral experimental finance. 2020;27:100352.
53. Modgil S, Gupta S, Stekelorum R, Laguir I. AI technologies and their impact on supply chain resilience during COVID-19. International Journal of Physical Distribution Logistics Management. 2022;52(2):13–49.
54. Yigitcanlar T, Kankanamge N, Regona M, Ruiz Maldonado A, Rowan B, Ryu A, et al. Artificial intelligence technologies and related urban planning and development concepts: How are they perceived and utilized in Australia? Journal of Open Innovation: Technology, Market, Complexity. 2020;6(4):187.
55. Kaur K, Rampersad G. Trust in driverless cars: Investigating key factors influencing the adoption of driverless cars. Journal of Engineering Technology Management. 2018;48:87–96.
56. Kowalczuk P. Consumer acceptance of smart speakers: a mixed methods approach. Journal of Research in Interactive Marketing. 2018;12(4):418–31.
57. Gansser OA, Reich CS. A new acceptance model for artificial intelligence with extensions to UTAUT2: An empirical study in three segments of application. Technology in Society. 2021;65:101535.
58. Weber RH. Internet of things: Privacy issues revisited. Computer Law Security Review. 2015;31(5):618–27.
59. Ding Y, Jin M, Li S, Feng D. Smart logistics based on the internet of things technology: an overview. International Journal of Logistics Research Applications. 2021;24(4):323–45.
60. Broeders D, Schrijvers E, van der Sloot B, Van Brakel R, de Hoog J, Ballin EH. Big Data and security policies: Towards a framework for regulating the phases of analytics and use of Big Data. Computer Law Security Review. 2017;33(3):309–23.
61. Montgomery KC. Youth and surveillance in the Facebook era: Policy interventions and social implications. Telecommunications Policy. 2015;39(9):771–86.
62. Gómez-Vallejo H, Uriel-Latorre B, Sande-Meijide M, Villamarín-Bello B, Pavon R, Fdez-Riverola F, et al. A case-based reasoning system for aiding detection and classification of nosocomial infections. Decision Support Systems. 2016;84:104–16.
63. Cumming D, Groh AP, Johan S. Same rules, different enforcement: Market abuse in Europe. Journal of International Financial Markets, Institutions Money. 2018;54:130–51.
64. Cybriwsky R. Changing patterns of urban public space: Observations and assessments from the Tokyo and New York metropolitan areas. Cities. 1999;16(4):223–31.
65. Anteby M, Chan CK. A self-fulfilling cycle of coercive surveillance: Workers' invisibility practices and managerial justification. Organization Science. 2018;29(2):247–63.
66. Nienaber A-M, Hofeditz M, Searle RH. Do we bank on regulation or reputation? A meta-analysis and meta-regression of organizational trust in the financial services sector. International journal of bank marketing. 2014;32(5):367–407.
67. De Keyser A, Bart Y, Gu X, Liu SQ, Robinson SG, Kannan P. Opportunities and challenges of using biometrics for business: Developing a research agenda. Journal of Business Research. 2021;136:52–62.
68. Li W, Yigitcanlar T, Liu A, Erol I. Mapping two decades of smart home research: A systematic scientometric analysis. Technological Forecasting Social Change. 2022;179:121676.
69. Elhai JD, Levine JC, Hall BJ. Anxiety about electronic data hacking: Predictors and relations with digital privacy protection behavior. Internet Research. 2017;27(3):631–49.
70. Li W, Zhou M, Dong H. CPT Model-Based Prediction of the Temporal and Spatial Distributions of Passenger Flow for Urban Rail Transit under Emergency Conditions. Journal of advanced transportation. 2020;2020(1):8850541.

Adoption and Implementation Trends of Vehicle-to-Everything (V2X) Technologies: A Comprehensive Bibliometric Analysis

Noor Islam Jasim, Saraswathy Shamini, Mohammed A. Al-Sharafi, Moamin A. Mahmoud, Muhammed Ibrahim, and Abba Hassan

Abstract The Vehicle-to-Everything (V2X) is a term used for denoting everything that the car can communicate with regarding its surroundings; it consists of vehicles, pedestrians, infrastructure, smart cities, and everything within its vicinity. V2X utilizes different types of communication technologies to communicate with other vehicles: infrastructure, networks and pedestrians, devices, electrical grid, homes, or the environment. The study aims to contribute by presenting a bibliometric analysis on global research trends related to V2X technology including publication outputs, co-authorship among authors; citation patterns and the countries' affiliation. We identified 6,958 articles meeting the inclusion criteria using the Scopus database among the total retrieved items (n = 8,432). China, the United States, and the United Kingdom are responsible for most of this growth in global publications (among 80 other countries), as shown by the results. Mohammad Saad Alam and Pavol Bauer were the most cited authors with the highest link strength in co-authorship networks, while Farukh Abbas and Mustafa Inci Stand out as highly-cited authors on V2X research. Prior authors summarize the information of 13 previous field studies, while our study provides a broad view only about V2X technology from Scopus database.

Keywords Vehicle to everything · V2X · Electric vehicle · Co-authorship · Co-occurrence

N. I. Jasim (✉) · M. Ibrahim · A. Hassan
College of Graduate Studies, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia
e-mail: noor.eng16@gmail.com

S. Shamini · M. A. Mahmoud
College of Computing and Informatics, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia

Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Kajang, Selangor, Malaysia

M. A. Al-Sharafi
IRC for Finance and Digital Economy, KFUPM Business School, King Fahd University of Petroleum & Minerals, Dhahran, Saudi Arabia

M. Ibrahim · A. Hassan
Faculty of Computing, Nigerian Army University, Biu, Borno State, Nigeria

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_18

1 Introduction

Vehicle-to-Everything (V2X) defines a technology that allows vehicles to communicate with pedestrians, automobiles, and the environment, like infrastructure or everything surrounding it within its perimeter [1]. V2X is an approach that employs many different communication technologies through which the vehicle connects with others around it, like vehicles, infrastructure, networks, pedestrians, devices, electrical grids etc. [2]. That is a huge change in the transport sector. V2X Integration aims to improve road safety and traffic efficiency, facilitate energy management, and provide better user experiences through communication between various subsystems of transportation systems. The potentiality of V2X would have transformative effects on future transportation. It is an essential factor influencing smarter transportation development and deployment. The future of transport will be highly influenced by its potential (V2X). Smarter transport systems are largely dependent on it. It can reduce accidents and make roads safer. V2X warns drivers about possible crashes, unsafe driving surfaces, or some sudden pedestrian motions [3]. Such adaptive supporting systems for higher traffic flows have been reported to develop with the help of V2X technology. Thus, optimizing utilization of car functionalities and green mobility [3]. In their research work, [4] studied the safety aspects of V2V and V2I in the context of vehicle-to-vehicle (V2V) communication protocols in cars, as well as those related to vehicle-to-infrastructure interactions within road networks. They further stated that real-time updates about congestion, alternate route suggestions, and coordination at intersections are made by V2X, hence facilitating efficient movement of traffic. This gives rise to new possibilities for energy management and sustainability on vehicle-to-household (V2H) and vehicle-smart grid (V2G) relationships.

The acceptance, adoption, use, and implementation of V2X must be studied alongside its innovative potential and opportunities. This study is intended to conduct a bibliometric analysis of the trend of adoption and implementation of V2X technologies. Understanding these trends will be useful in understanding state-of-the-art technology, trends, and challenges for V2X. Policymakers would enable informed strategies based on their specific needs; industries could make investment decisions with proper guidance, and stakeholders could make appropriate informed financial decisions. Researchers would unlock academia's potential and redirect efforts toward more promising directions. Bibliometric analysis has been cited by Donthu et al. [5] as a popular but rigorous scientific data exploration method where large volumes of data are under scrutiny. It lends insights into the evolutionary patterns peculiar to a particular field or area within it. It is powerful because it uses statistical procedures such as publication trends, citation patterns, and keyword co-occurrences, among others, to help evaluate patterns in scientific discourse. If this approach is used, along with visual tools such as VOSviewer, to map V2X literature, this study will be able to determine emerging adoption and implementation trends of V2X technologies.

Nonetheless, the broad language of keyword co-occurrences suggested that V2X literature was skewed to technical aspects. For example, 5G mobile communication systems have a total link strength of 4155, and then followed by electric vehicle

(EV), then battery-electric-vehicle (BEV). Indeed, this indicates less study regarding widespread acceptance, adoption, or usage. However, a recent study by Alalewi et al. [6] on 5G V2X communications has identified that in practice this goal reaches sub-1 ms latency performance is difficult to sustain for high-mobility vehicular scenarios. Empirical tests in the urban setting observed spikes of up to 20 ms latency during inter-base station handovers that could lead catastrophic occurrences, such as emergency braking situations.

V2X adoption is still hindered by major concerns around security and privacy. A foundational survey of security challenges in vehicular ad hoc networks (VANET) was made by Hasrouny et al. [7]. Additionally, they categorized threat types into identity-based, such as Sybil attacks, location-based threats, like GPS spoofing, and data attacks, for example, replay. They stress that V2X is highly distributed and operates in a high-speed environment; traditional security solutions are suboptimal for this use case. Privacy was also studied in V2X communications by Rathee et al. [8] finding that cars sending status messages can be recognized and tracked which endangers user privacy. These authors suggested a pseudonym-based scheme, conceding to high computational overhead (and not improved in the past 14 years), illustrating the ongoing trade-off between security/privacy and performance.

Economic factors play a massive role in V2X adaption. Massive infrastructure investments, particularly for RSUs and 5G networks, that will be needed represent a big problem. Gschwendtner et al. [9] gives an overview of the trial configurations and potential to defer grid reinforcement, noting vehicle-to-customer (V2C) services are more common than transmission-level. Social aspects; Incentives and operations should be tailored, Regulation; needs to avoid double taxation, Market participation/ V2X standards. It could make the company more money and reduce exposure from changing driving habits and charging behavior. Also, mobility cycles are short 1–2 years) compared to car product lives (5–7) years; the automotive manufacturers will be risking themselves on a specific V2X technology that may become outdated [10].

Information governance is another vital regulatory challenge. V2X is data-intensive with extensive data sharing; this poses questions related to ownership, access, and monetization aspects essential for the sustainable functioning of the system powered. In a study by Islami et al. [11]. These questions were approached from the perspective of GDPR, and so they argued that V2X data is highly sensitive, such as driving habits or location, which has to be anonymized before being shared. They also caution that ambiguous data rights could undermine consumer confidence or enable predatory data practices, potentially inhibiting adoption.

This bibliometric investigation on V2X adoption, factors, and challenges in the literature revealed a complex landscape. Additionally, due to technical challenges such as latency, those tasks cannot be safety–critical in a V2X environment yet. Quality of data for security and privacy risks, eroding public confidence. The economic obstacles, for instance infrastructure costs that are too high, enables little deployment. The grim reality of inconsistent regulations and unresolved issues around data ownership and liability only compound the confusion. Collectively, those challenges form a mosaic of enormous potential for technology but also are an aligned list of barriers that must be addressed in multi-disciplinary approaches to realize widespread adoption. Since there is needed to mitigate these challenges and enhance V2X technologies and adoption, it brings the research question of this study. In this paper, we intend to answer these questions through the exploration of:

1. Who are the main co-authors, their institutions, citations, research trends, and contributing countries towards V2X studies?
2. Which keywords are most frequently used in V2X studies?

The subsequent sections of the study are organized as follows: Sect. 2 focuses mostly on the employed approach, i.e., search strategy, with reference to inclusion/exclusion criteria and data analysis procedures. Section 3 presents the results analysis, particularly the publication trends, country contributions, co-authorship patterns, and key themes through visually appealing figures and tables. Section 4 elaborates on the analysis and critical review of research findings and offers an engaging discussion with a special focus on transversal topics as well as interdisciplinary perspectives and global dimensions. Section 6 introduces the impact of V2X technologies. Section 7 concludes by highlighting insights and implications, as well as opportunities for future research.

2 Methodology

2.1 Data Sources Scopus

This bibliometric study intends to analyze publication trends using the V2X literature indexed in the Scopus database. Bibliometric analysis—a systematic and methodical examination of data from academic literature databases to provide an overview of the global research trends in a specific field [4, 12, 13]. This approach not only identifies key publications, influential authors, and leading institutions but also maps the evolution of the research domain over time [14, 15]. By examining citation patterns, co-authorship networks, and keyword occurrences, bibliometric analysis offers valuable insights into the structure and dynamics of the field, enabling the identification of emerging themes, research gaps, and potential areas for future investigation. By utilizing bibliometric techniques, this study will identify key patterns in publication activity, including the most influential authors, institutions, and countries contributing to the V2X research domain. Additionally, the analysis will reveal the evolution of research themes, the most frequently cited papers, and emerging areas of interest within the V2X field. The insights gained from this analysis will highlight the current state of V2X research and help identify potential gaps and opportunities for future studies.

2.2 Search Strategy and Search String

A systematic search strategy was employed to identify relevant literature on V2X technology. The search terms were carefully selected to encompass various aspects of V2X technology and include bibliometric studies. The final search string used in the database was: ("Vehicle-to-Everything" OR "V2X" OR "Vehicle-to-Infrastructure" OR "Vehicle-to-Vehicle" OR "Vehicle-to-Network" OR "Vehicle-to-Pedestrian" OR "Vehicle-to-Device" OR "Vehicle-to-Grid" OR "Vehicle-to-Home" OR "Vehicle-to-Road") AND ("adoption" OR "implementation" OR "usage" OR "utilization" OR "acceptance"). This comprehensive approach ensured that relevant literature covering different facets of V2X technology and its adoption was captured and included in the analysis.

2.3 Inclusion and Exclusion Criteria

The initial search yielded 8,432 documents. To refine the analysis, we applied additional criteria: the studies had to be in English, published in the period of 2014–2024, that are relevant to the domain of the study. This filtering process led to the exclusion of 1,474 papers that did not align with the study's objectives (see Fig. 1).

3 Result Analysis

A collection of 6,958 articles was exported from Microsoft Excel (2016) to VOSviewer (version 1.6.20), including author citation information, bibliographic details, and author keywords. VOSviewer is a software that specializes in generating and visualizing bibliometric maps .

3.1 Trend in Publication Analysis

Based on the analysis, the result shows the current trends in V2X technology and how it is advancing rapidly, supported by substantial research contributions from various conferences and journals, as shown in Fig. 2. "IEEE Access" stands out in the V2X field with the highest citation of 4542 and 3947.86 link strength in total. Its high count of citation and link strength underscore its crucial role in influencing V2X research, offering valuable insights into technical innovations and implementation strategies. "Applied Energy" is another prominent journal with citations and link strengths reaching up to 1470 and 1307.02, respectively, emphasizing the intersection of V2X with energy systems and sustainability. IEEE conferences have been

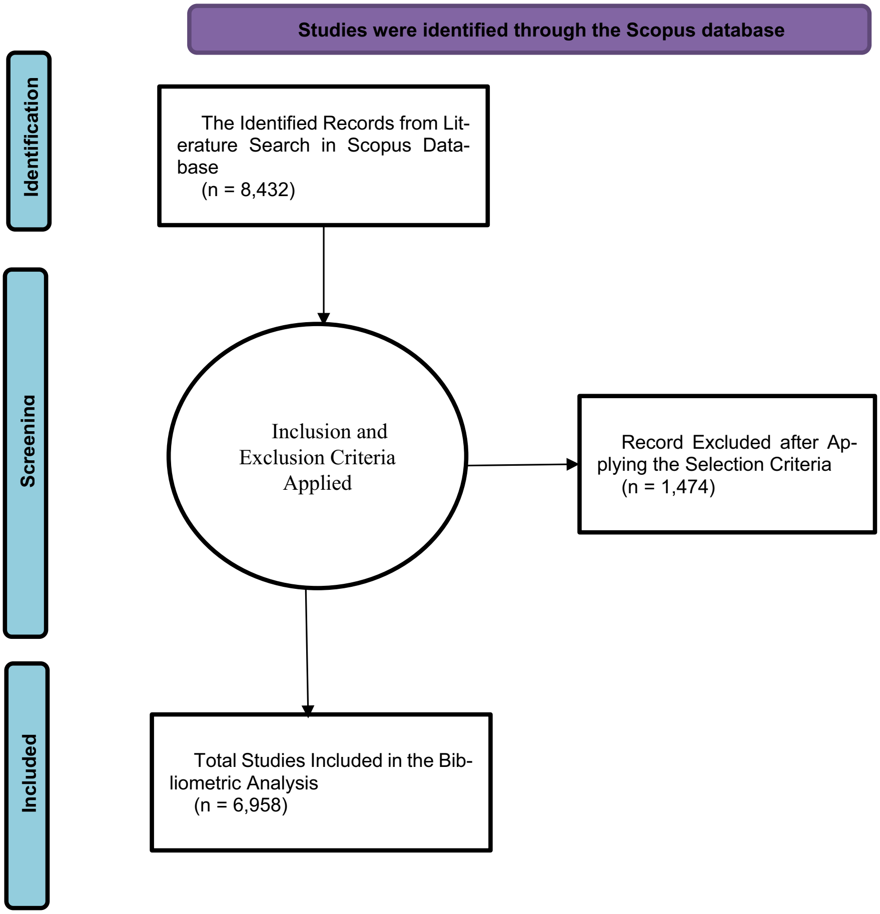


Fig. 1 Prisma flow diagram

instrumental in advancing V2X research. The 2010 IEEE Vehicle Power and Propulsion Conference, with 21 citations and a link strength of 20.65, marks early interest in V2X technologies. This momentum continued with the 2011 IEEE Vehicle Power and Propulsion Conference, which garnered 20 citations and a link strength of 18.86. The 2012 IEEE Power and Energy Society General Meeting, notable with 26 citations and a link strength of 25.65, highlights the integration of V2X technology with power and energy systems. The 2016 IEEE Vehicular Networking Conference (VNC) and the 2018 International Conference on Intelligent Transportation Systems (ITSC) are also significant, with 21 and 24 citations, respectively, and link strengths of 18.86 and 23.31. The 2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring) contributes notably with 20 citations and a link strength of 20.00. Accident Analysis and Prevention is a leading journal in V2X research, with multiple entries having citations ranging from 43 to 136 and link strengths from 30.27 to 97.68.

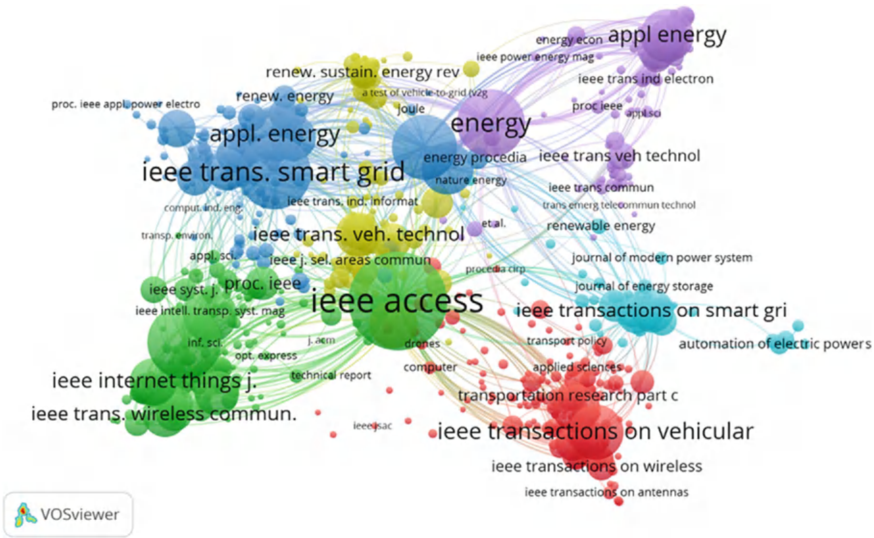


Fig. 2 Research sources and current trend

Its focus on safety and prevention aligns with V2X technology's goal of reducing traffic accidents and enhancing road safety. "ACM Computing Surveys" and "ACM SIGCOMM Computer Communication Review" have made significant contributions, as evidenced by their robust citation and link strength metrics. These journals' interdisciplinary nature highlights the diverse applications of V2X technology, from soft computing techniques to broader scientific inquiries.

3.2 Countries Analysis

Based on the results of the analysis, the studies were conducted in 210 countries. A criterion was set that included a country with a minimum of 5 documents; 80 countries met the criteria. The results indicate a significant knowledge regarding the countries with the highest documents, citations, and total link strength in the research field. China leads in the number of documents, with a total of 1,552 publications. This indicates a substantial research output in the field. Additionally, China has 25,888 citations, reflecting the significant impact and recognition of its research. The country's high total link strength of 786 suggests robust international collaborations, making China a pivotal player in global research networks. The United States follows closely with 1,210 documents and an impressive 44,582 citations. This high citation count demonstrates the profound influence and quality of US research. The United States also boasts a total link strength of 635, indicating strong international research connections and collaborations, further amplifying the global reach

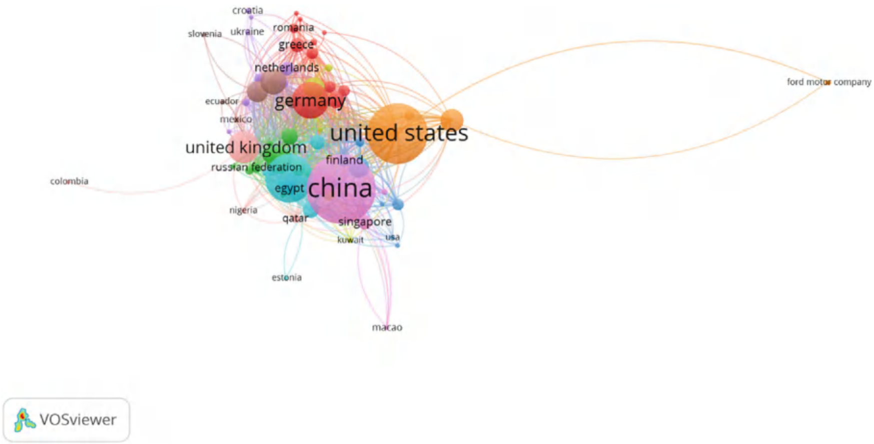


Fig. 3 Countries of the studies

of its research outputs. India also shows notable participation, with 837 documents and 8,130 citations. This substantial volume of publications and citations highlights India's active engagement in the research area. India's total link strength of 284 suggests good international cooperation, contributing to the visibility and impact of its research. The United Kingdom stands out for its high citation count of 11,103, derived from 406 documents. This indicates that UK research outputs are of high quality and extensively cited by other researchers. The UK also has the highest total link strength of 426, reflecting the strongest international research collaborations and networks. This extensive networking enhances the dissemination and influence of UK research globally. Germany and Canada also make significant contributions. Germany has 520 documents and 8,827 citations, while Canada has 276 documents and 8,393 citations. Both countries exhibit substantial research output and impact, with Germany's total link strength at 217 and Canada's at 272. Figure 3 shows the countries that the studies were conducted in.

3.3 Co-authorship Analysis

Based on the bibliometric analysis of the result, 20,978 authors were found, and the criteria set 5 as the minimum documents and five minimum citations, 297 meet the criteria as shown in Fig. 4. On co-authorship, focusing on authors with the highest documents, citations, and total link strength, several prominent researchers emerge. Firstly, Mohammad Saad Alam stands out with nine documents, accumulating an impressive 742 citations and a total link strength of 3. His work is influential in the field, and his extensive citation count underscores the significant impact of his research. Next, Pavol Bauer, with 8 documents, garners 391 citations and a total link strength of 7. Bauer's research contributions are well-regarded, and his strong link strength indicates his pivotal role in collaborative research networks. Alessandro Bazzi, with 11 documents, 645 citations, and the highest total link strength of 16, is another key figure. Bazzi's high link strength reflects his extensive collaborations and the importance of his research within the academic community. Piyush Bhagdikar, although having eight documents and 16 citations, exhibits a remarkably high total link strength of 26, suggesting a significant role in collaborative efforts despite the lower citation count. Sukru Yaren Gelbal and Omprakash Kaiwartya, both with total link strengths of 13 and 1, respectively, have a modest number of documents and citations, but their high link strengths highlight their collaborative influence. Xiaojiang Du, having authored six documents, while the citations are 71 together with seven strength links in total, also indicates strong collaborative ties, contributing effectively to his research community. Dusit Niyato, with seven documents, the citations are 435 18 strength links, showcases a blend of prolific studies output, substantial citation impact, and extensive collaboration. Xuemin Shen, with a remarkable 18 documents, 1224 citations, and a total link strength of 25, stands out as a leading researcher with substantial contributions and strong collaborative networks. Lastly, Salman Habib, with five documents and 403 citations but a no strength link, indicates a high impact in terms of citations, despite limited collaborative connections.

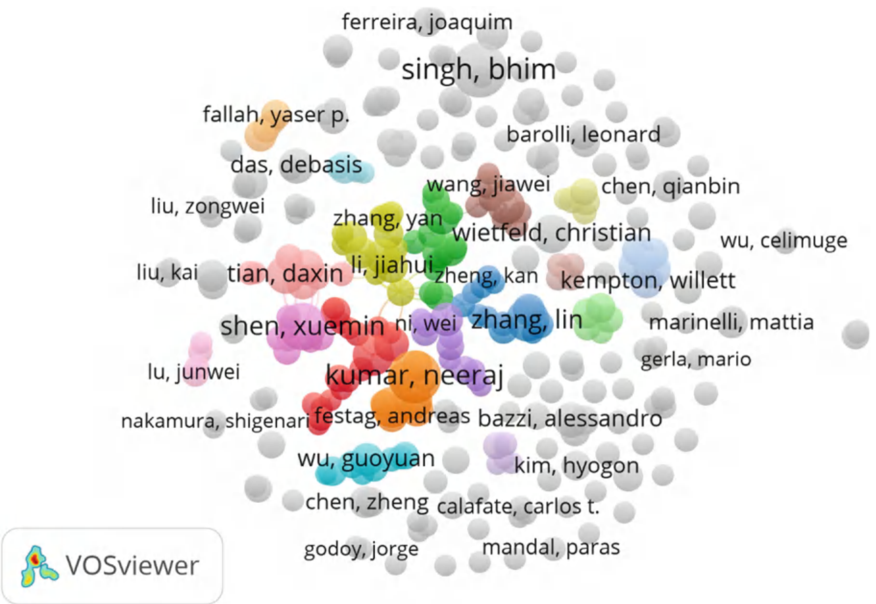


Fig. 4 Co-authorship

3.4 Co-occurrence Keywords Analysis

The analysis of bibliometric data reveals several prominent trends and focal points within the research landscape, as shown in Fig. 5. One noteworthy trend is the significant emphasis on 5G mobile communication systems, as indicated by its substantial total link strength of 4155. This suggests a strong emphasis on advancing mobile communication technologies, likely in response to the increasing demand for faster and more reliable wireless communication networks. Another significant focus is electric vehicles (EVs), highlighted by frequent keywords such as "electric vehicle," "battery electric vehicle," and "electric vehicle chargers," which show a strong interest in sustainable transportation and EV infrastructure development. This aligns with global efforts to lower greenhouse gas emissions and address climate change by adopting more sustainable transportation methods. Energy management is another prominent research area, as reflected by keywords like "energy management systems," "energy management," and "energy optimization," emphasizing the importance of enhancing energy efficiency and sustainability across various sectors, including transportation, manufacturing, and building management. This focus on energy management supports broader goals of reducing energy use and carbon emissions. Additionally, artificial intelligence (AI) is noted as a crucial technology advancing innovation across different domains, with the keyword "artificial intelligence" appearing frequently, indicating a growing interest in AI applications like machine learning and data analytics in fields such as healthcare, finance, and autonomous systems. The emphasis on AI highlights its potential to revolutionize industries and improve efficiency and decision-making. Additionally, there is a significant concern for environmental impact in the research landscape, with keywords like "environmental impact," "greenhouse gas emissions," and "global warming" indicating a strong emphasis on sustainability and environmental conservation. This suggests an increasing awareness of the need to address environmental challenges and develop sustainable solutions to mitigate the effects of climate change. The emphasis on environmental impact aligns with global initiatives to advance sustainable development and minimize carbon emissions. In summary, the bibliometric analysis underscores significant trends and focal points in research, encompassing developments in mobile communication systems, the shift toward electric vehicles, endeavors to enhance energy management and sustainability, the impact of artificial intelligence on innovation, and the emphasis on environmental impact and conservation. These trends mirror evolving research priorities and the worldwide endeavor to tackle intricate challenges and forge a more sustainable future.

3.5 Citation Analysis

The bibliometric analysis offers valuable insights into the research output and influence of various authors in the field, as indicated in Fig. 6. Mustafa NCI stands out with

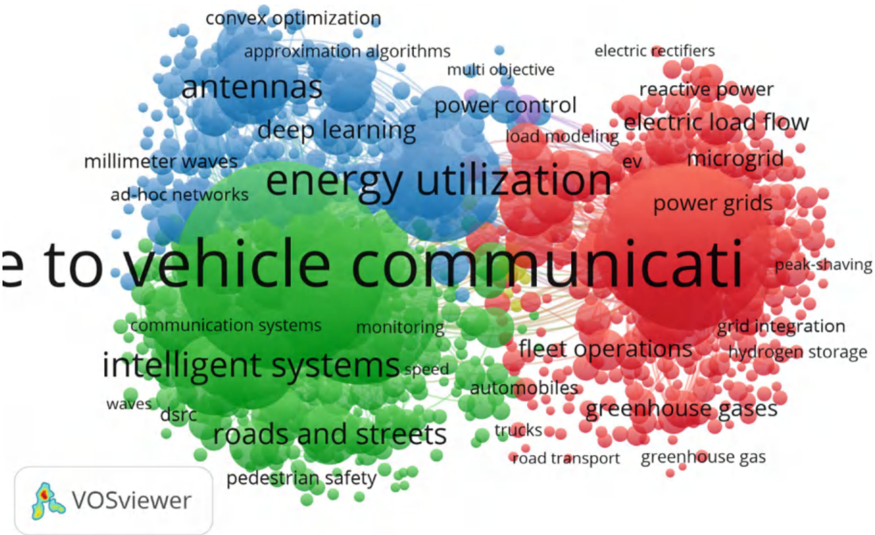


Fig. 5 Co-occurrence of keywords

three documents and 113 citations, indicating significant contributions. However, their total link strength of 26 suggests a moderate level of connectedness within the research network. Luis F. Abanto-Leon has also made notable contributions with three documents and 48 citations, although their total link strength is relatively low at 3, indicating limited connectedness. Fakhar Abbas has four documents and 29 citations, with a total link strength of 13, reflecting a moderate level of connectedness similar to Farukh Abbas is a standout author with three documents and the highest citation count of 346, indicating a substantial impact. Their total link strength of 56 suggests a well-connected body of work within the research community. Anuj Abraham has contributed significantly with five documents and 20 citations, although their total link strength is 7, indicating moderate connectedness. João L. Afonso, with three documents and 103 citations, has a total link strength of 10, suggesting a moderate level of connectedness within the research network. Shubhani Aggarwal has three documents and 96 citations, with a total link strength of 27, indicating a well-connected body of work. Vaneet Aggarwal has three documents and 53 citations, with a total link strength of 9, indicating moderate connectedness. Jamshid Aghaei has made significant contributions with five documents and 107 citations, although their total link strength of 23 suggests a moderate level of connectedness.

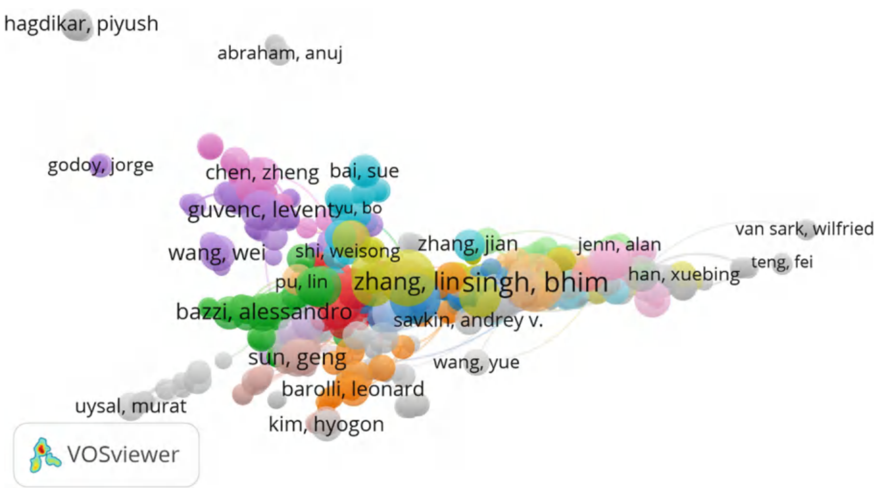


Fig. 6 Citation analysis

4 Future Directions and Emerging Themes

This study highlights current stage of V2X research as up to publication date, and also serves as a compass that points toward V2X future directions and emerging themes. By examining publication trends, citation patterns, keyword co-occurrences, and author collaborations, the study identified the trajectory of V2X technology and areas leading to groundbreaking research (Table 1).

5 Discussion

This study exploits the Scopus database to investigate current trends in V2X technologies, adoption, and usage. It is apparent technology holds immense potential for transforming transportation systems to be more intelligent. However, a multifaceted paradigm brimming with both opportunities and challenges has been revealed. The result analysis highlights technical advancements as one of the most striking trends for the V2X. Significant research is leaned on 5G communication systems, inferring that V2X requires a high-speed and reliable data exchange to function accurately. Furthermore, EV prominence reflects a vital collaboration between V2X and sustainable transport solutions. V2X technologies optimize energy management through charging infrastructures and utilization of overall energy efficiency.

In spite of this, the trail in the direction of V2X adoption revealed a significant challenge that requires attention. For example, security and privacy concerns are a significant adoption barrier. V2X technologies rely on data exchange to function

Table 1 Future directions and emerging themes

Themes	Challenges	Directions
User behavior and adoption	Low user awareness, privacy concerns	Explore factors influencing user acceptance and integration of V2X technologies Understand user trust, privacy concerns, and willingness to adopt V2X
Infrastructure investment	High cost of V2X infrastructure deployment	Explore cost-effective solutions, public–private partnerships
Policy and regulation	Lack of clear regulations	Develop robust regulations for data privacy, cybersecurity, and standardization to facilitate V2X implementation
Global implementation	Limited research in developing regions	Expand research to identify challenges and opportunities in developing regions
Interdisciplinary approach	Limited research in other fields	Foster collaboration across different fields, particularly automotive, technology, urban planning, and social science sectors, to address the complex challenges of V2X technology
Sustainability	GHG emission, traffic congestion, noise pollution, energy consumption	Explore how V2X can contribute to energy efficiency, reduced traffic congestion, and a more sustainable transportation system

accurately. This leads to the vast amount of data exchange between vehicles, pedestrians, roads, and other infrastructures that raises anxieties about potential breaches, ownership, and misuse. Clear and standard solutions are needed to ensure user trust and mitigate these worries. The analysis also highlights economic factors as a barrier associated with V2X infrastructure deployment. To have a full stack V2X, there is a need to upgrade the roads and its network, traffic signals, and other substructures to communicate, hence requiring significant funding. There is a need to devise a cost-effective solution to overcome this obstacle.

Inadequate regulations are another challenge that deters V2X implementation. A unified communication is vital to V2X; standard protocols for data privacy, cybersecurity, and interoperability are required between vehicles and infrastructure from different manufacturers. To facilitate widespread adoption, vigorous V2X regulatory frameworks that foster a secure and reliable ecosystem should be devised. The study underscores a slant of interdisciplinary collaboration and global cooperation. V2X would not be limited to technology and ICT because it is a multifaceted paradigm that requires expertise from automotive engineering, computing, social science, and urban and regional planning. These fields require vigorous collaboration to address the technical, social, and economic challenges associated with V2X adoption. Similarly, user behavior is a vital factor that influences widespread adoption. Understanding determinants such as user trust and privacy concerns will be crucial for the successful implementation and adoption of V2X technology.

This study aligns with previous research by Mahi et al. [16] Who identified similar challenges such as security and privacy concerns, and the need for clear regulations. This study extends this by highlighting the increasing focus on user behavior and the potential of V2X for sustainability. The findings suggest the need for policy interventions to incentivize V2X adoption. Tax breaks on V2X-enabled vehicles, subsidies for infrastructure upgrades, and robust data privacy regulations are some potential policy directions. The study is limited to the Scopus database, and alternative databases might reveal additional insights. Additionally, the chosen search terms could influence the results.

6 Implication

Both researchers, V2X policymakers as well as the industries themselves will find this study significant for tailored service destined to customers. It illustrates how V2X technology is rapidly becoming a critical part of state-of-the-art transport, requiring modifications on both sides of this vehicle-city infrastructure. Effective regulatory frameworks must be implemented that facilitate V2X deployment, covering issues relating to data privacy, cybersecurity, and standardization. This calls for collaboration and innovation between different streams of knowledge, which is critical in managing the complexity associated with V2X technology. For V2X communication to be beneficial, infrastructure investments are required—everything from updating roads and signals to communications networks. This research delivers insights for the automotive sector as well as tech players and highlights burgeoning V2X vehicle service opportunities. The research spans a wide range of topics within V2X, and it was seen that many focus on the technical aspect of it instead of the behavior of the technology.

7 Conclusion

This study provides a detailed look at the current trends in V2X research by analyzing a collection of articles from the Scopus database. The findings show a significant interest and development in this field over the past decade, and it's expected to keep growing. Countries like China, the USA, and India are leading the charge, fostering strong international collaborations. The study also highlights global trends and regional differences in V2X research and implementation, advocating for international collaboration to share best practices and advance V2X technology worldwide. This opens the door for researchers from other parts of the world, particularly in Africa and developing regions, to get involved and contribute to this exciting area of study.

Acknowledgements This work was supported by Dato' Low Tuck Kwong International Energy Transition Grant under the project code 202202003ETG.

References

1. A. W. Thompson and Y. Perez, "Vehicle-to-Everything (V2X) energy services, value streams, and regulatory policy implications," *Energy Policy*, vol. 137, p. 111136, Feb. 2020, https://doi.org/10.1016/j.enpol.2019.111136.
2. M. A. Rehman, M. Numan, H. Tahir, U. Rahman, M. W. Khan, and M. Z. Iftikhar, "A comprehensive overview of vehicle to everything (V2X) technology for sustainable EV adoption," *J. Energy Storage*, vol. 74, p. 109304, Dec. 2023, https://doi.org/10.1016/j.est.2023.109304.
3. K. R. Dhinesh and A. Rammohan, "Revolutionizing Intelligent Transportation Systems with Cellular Vehicle-to-Everything (C-V2X) technology: Current trends, use cases, emerging technologies, standardization bodies, industry analytics and future directions," *Veh. Commun.*, vol. 43, p. 100638, Oct. 2023, https://doi.org/10.1016/j.vehcom.2023.100638.
4. S. Adnan Yusuf, A. Khan, and R. Souissi, "Vehicle-to-everything (V2X) in the autonomous vehicles domain – A technical review of communication, sensor, and AI technologies for road user safety," *Transportation Research Interdisciplinary Perspectives*, vol. 23. Elsevier, p. 100980, Jan. 2024. https://doi.org/10.1016/j.trip.2023.100980.
5. N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, no. May, pp. 285–296, Sep. 2021, https://doi.org/10.1016/j.jbusres.2021.04.070.
6. A. Alalewi, I. Dayoub, and S. Cherkaoui, "On 5G-V2X Use Cases and Enabling Technologies: A Comprehensive Survey," *IEEE Access*, vol. 9. Institute of Electrical and Electronics Engineers Inc., pp. 107710–107737, 2021. https://doi.org/10.1109/ACCESS.2021.3100472.
7. H. Hasrouny, A. E. Samhat, C. Bassil, and A. Laouiti, "VANet security challenges and solutions: A survey," *Veh. Commun.*, vol. 7, pp. 7–20, Jan. 2017, https://doi.org/10.1016/J.VEHCOM.2017.01.002.
8. G. Rathee, A. Sharma, R. Iqbal, M. Aloqaily, N. Jaglan, and R. Kumar, "A Blockchain Framework for Securing Connected and Autonomous Vehicles," *Sensors*, vol. 19, pp. 1–15, Jul. 2019, https://doi.org/10.3390/s19143165.
9. C. Gschwendtner, S. R. Sinsel, and A. Stephan, "Vehicle-to-X (V2X) implementation: An overview of predominate trial configurations and technical, social and regulatory challenges," *Renew. Sustain. Energy Rev.*, vol. 145, p. 110977, Jul. 2021, https://doi.org/10.1016/J.RSER.2021.110977.
10. N. Lu, N. Cheng, N. Zhang, X. Shen, and J. Mark, "Connected Vehicles: Solutions and Challenges," *Internet Things Journal, IEEE*, vol. 1, pp. 289–299, Aug. 2014, https://doi.org/10.1109/JIOT.2014.2327587.
11. L. Islami, A. Kitkowska, and S. Fischer-Hübner, "Inter-regional Lens on the Privacy Preferences of Drivers for ITS and Future VANETs," vol. 24, pp. 1–20, May 2024, https://doi.org/10.1145/3613904.3641997.
12. A. O. Baarimah *et al.*, "A bibliometric analysis and review of building information modelling for post-disaster reconstruction," *Sustain.*, vol. 14, no. 1, pp. 1–20, Dec. 2022, https://doi.org/10.3390/su14010393.
13. H. Ghaleb, H. H. Alhajlah, A. A. Bin Abdullah, M. A. Kassem, and M. A. Al-Sharafi, "A Scientometric Analysis and Systematic Literature Review for Construction Project Complexity," *Buildings*, vol. 12, no. 4, 2022, https://doi.org/10.3390/buildings12040482.
14. M. Ibrahim, M. A. Mahmoud, M. A. Al-Sharafi, A. Hassan, N. A. Iahad, and Saraswathy Shamini Gunasekaran, "Smart Home Technology Research Landscape: A Bibliometric Analysis of Scientific Publications (2014–2024)," *Stud. Comput. Intell.*, vol. 1161, pp. 97–115, 2024, https://doi.org/10.1007/978-3-031-61463-7_6.

15. A. Hassan, M. A. Mahmoud, M. A. Al-Sharafi, M. Ibrahim, N. A. Iahad, and S. S. Gunasekaran, “Bibliometric Analysis of Electric Vehicle Adoption and Environmental Sustainability Research (2014–2024): Trends and Implications,” *Stud. Comput. Intell.*, vol. 1161, pp. 43–58, 2024, https://doi.org/10.1007/978-3-031-61463-7_3.
16. M. J. N. Mahi *et al.*, “A Review on VANET Research: Perspective of Recent Emerging Technologies,” *IEEE Access*, vol. 10, pp. 65760–65783, 2022, https://doi.org/10.1109/ACCESS.2022.3183605.

Artificial Intelligence Methods and Image Recognition Techniques in Ophthalmic Robotic Surgery: A Review

Mukesh Madanan, Saraswathy Shamini Gunasekaran, Moamin A. Mahmoud, Jaspaljeet Singh Dhillon, Salama Mostafa, and Nazirul Nazrin Shahrol Nidzam

Abstract Use of robots in ophthalmic surgery to improve control of movement, vibrations termination, vision, and distance detection, makes it easier to perform complex eye surgeries. The use of Artificial Intelligence (AI) in ophthalmology is finding its way in ocular surgery that utilises surgical robots with specialized microscopes and visualization systems for surgical process. Proper visualization process in ocular robotic surgery can be achieved by introducing image recognition techniques in the robots. Image recognition for surgical instrument detection and segmentation of the surgical areas can be performed effectively through using AI based approaches such as deep learning techniques. In this article, the use of robots and ophthalmology tele surgery is examined, along with the obstacles impeding its quick development. Performance of various deep learning-based image recognition approaches are compared using metrics such as Accuracy (95%), Recall (91%), Error (0.05), and Precision (90%). Through this study, it is suggested that AI technologies are needed to reduce resource scheduling problems in tele-surgery and improve reliability and efficiency of ocular surgery.

Keywords Deep learning · SVM · CNN · Robotic surgery · Artificial intelligence · Ophthalmology · Tele surgery · Surgical tool detection

S. S. Gunasekaran · M. A. Mahmoud · J. S. Dhillon
College of Computing and Informatics, Universiti Tenaga Nasional, Selangor, Malaysia

Institute of Informatics and Computing in Energy, Universiti Tenaga Nasional, Selangor, Malaysia

M. Madanan (✉) · N. N. S. Nidzam
College of Graduate Studies, Universiti Tenaga Nasional, Selangor, Malaysia
e-mail: mukeshmvm@gmail.com

S. Mostafa
Faculty of Computing, Universiti Tun Hussein Onn, Johor, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_19

1 Introduction

The introduction of robotics into the medical sector has been considered as an achievement for the medical practitioners. Although the use of robotics has been used in diagnosis and treatment of several diseases, the usage in surgery is still a dark patch to a certain extend when it comes to autonomy. Surgery is the most sought out area of medical practitioners and robotics have found its way in semi-autonomous surgery and partial human–robot surgery. The image recognition feature is the most vital feature of robotic surgery that recognizes the area of the surgery where the robot has to perform the surgical process.

1.1 Rationale

Over the past three decades, computerized surgical systems have undergone significant growth and integration. When compared to robotic surgery in general, robotic eye surgery has a very young history. While other surgical professions have gained extensive clinical experience using robotic surgery on patients, ophthalmic surgeons have focused on showing the viability of certain eye operations using a robot in a laboratory setting [1]. Due to the developments in the robotic system, tele-surgery is a revolutionary new method of providing surgical treatment [2]. In tele-surgery, the surgeon operates the machine while seated at a console far away from the patient with the use of advanced wireless networking techniques and robotics are required for this surgical system to carry out the procedure [3]. The primary goal of tele surgery is to reduce surgeons' travel to patients' area and their companions [4].

Artificial intelligence (AI) was initially applied to the domains of medicine and surgery, especially ophthalmology, to identify a variety of illnesses of the eye using deep learning (DL), machine learning (ML) and other image recognition techniques [5]. The detection of eye illnesses, the mapping of the cornea's topography, and the computation of intraocular lenses have all been validated using AI, ML, and DL. Retinal damage from diabetes, age-associated macular degeneration (AMD), and cataracts are the top three root causes of permanent vision loss in the world that could be detected using AI and image recognition [6]. Ophthalmology is witnessing the emergence of advanced AI applications that will be employed in clinical investigations, treatment and surgical instructions, and continuous patient care. Robotic technology is improving surgical movement controls, tool tremor controls, surgical area visualization, and distance sensing in ophthalmic surgery [7]. For use in ophthalmology procedures, automated devices and models must meet speed and velocity requirements, while preserving path-tracking accuracy is necessary to meet the time delay threshold [8]. Medical and surgical specializations with specific requirements are found in the area of ophthalmology [9]. Robotics integration in eye surgery is challenging due to the requirement for surgical microscopes, as well as viewing systems and parameters. There haven't been much research connecting robotics with tele-surgery in ophthalmology, but a study of the present state of the existing fields might help its progress [10]. An accelerometer may quickly eliminate the operator's physiological tremor while robotic arms reduce the natural constraints of human wrists to enable greater accuracy and efficiency actions [11]. This method has drawbacks including a higher chance of human errors, mechanical failure, improper use of electrical power, and prolonged operation timings [12].

Robot-assisted vitreoretinal surgery was accurate and less traumatic to the tissues than manual surgery, although it was slower [12]. The poor availability of surgical robots in healthcare systems makes the access challenging and time-consuming, and their exorbitant cost discourages many investigations and trials on ocular surgery [13]. There is a need for particular robotic ocular surgery prototypes that can improve accuracy as well as technologies like tactile internet and artificial intelligence to alleviate resource scheduling issues in tele surgery [14]. The area of robotic surgery might advance by having more clinical training programs [15]. The use of AI in robotic surgery involves image recognition combined with several other operations or functions such as movement adjustment, friction monitoring etc. This requires the employment of several machine learning and deep learning algorithms to achieve the required results. Currently the ophthalmic surgery is performed semi-autonomously where the surgeon and the robot behave as master–slave. To achieve full autonomy in ophthalmic surgery, the robot must focus on image recognition and take the feedback from it and then position the surgical tool automatically on the surgical area, which is currently positioned by the surgeon.

The paper focuses on studying the introduction of robots in ophthalmic surgery for performing ophthalmic surgeries more accurately. Robotics assisted surgery in different parts of eye is then focused. Following that, use of tele surgery in existing robotic eye surgery is investigated. Finally, the use of various AI techniques for performing image recognition in robotic assisted eye surgery is investigated. Image recognition can help in instrument detection and segmentations for effective surgery.

2 Use of Robotics in Surgical Field

Robotic Surgery, commonly referred to as robot-assisted surgery, gives surgeons the ability to perform a variety of challenging procedures with more accuracy, mobility, and monitoring than previously possible with conventional methods.

Stoianovici [16] demonstrated the robotic system to facilitate and improve manufacturing. Robotics are currently chosen to improve the delivery of quality health care (Table 1).

The articles evaluated above are related to the use of robotics in different surgical areas of human body. Da Vinci surgical system, boasts of independent modular arms that are "quick and easy to set up" and promises to be more flexible and versatile. The surgery using Da Vinci surgical system may be performed while standing or sitting and can fit in practically in any operating room because of its small size.

Table 1 Analysis of use of robots in medical field

Contributor(s)	Area of research	Focus of research
Hillel et al. [17]	Used in laryngeal surgery	Improve surgical precision
Hanly and Talamini [18]	Expand the scope of minimally invasive surgery	Improves vision and repair
Senapati and Advincula [19]	Advanced laparoscopy	Improve telemedicine and telepresence surgery
Thiel and Winfield [20]	Urology using da Vinci surgical system	Reduce invasiveness and improve clinical outcomes
Sim et al. [21]	Analysed a camera arm and movable arms with surgical tools	Controls the arms from a console
Xiong et al. [22]	Robotic surgery teleoperation	Finer flexibility and improved mobility during minimal invasive surgery
Lehman et al. [23]	For vision and task support in laparoscopic operations	Minimising patient stress and speeding up recovery
Zemmar et al. [24]	Reduce viral infection and enhance patient care	Spread of diseases in human body
Camps [25]	Feasibility, safety, and advantages of RMIS	Reduces invasiveness and provides better visualization and dexterity

3 Traditional Manual Surgery in Ophthalmology

Ophthalmology includes diagnosis of the eyes, treating eye diseases and performing ophthalmic surgery. With several treatments being developed for the treatment of eye issues, it is still a genre of surgery that needs to be developed using modern techniques and tools. The Fig. 1 represents the human surgeon utilizes an optical microscope to provide visual input while controlling the surgical tools in conventional surgery.

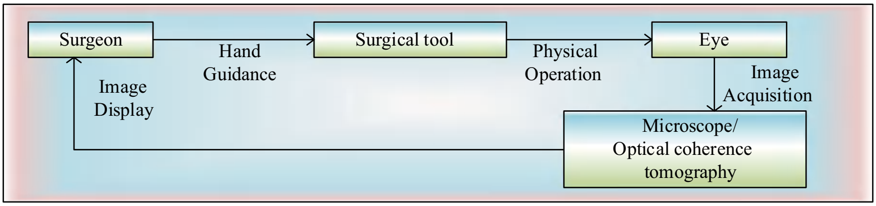


Fig. 1 The tactile navigation system's process flow in manual surgery [24]

4 Robotics Assisted Surgery in Ophthalmology

Ophthalmology has only recently initiated to employ robotic surgical assistance. The da Vinci Surgical System employs tele-manipulation to carry out ex-vivo ocular operations and to correct pterygium despite not being especially created for ophthalmic treatments. Figure 2 shows how a surgical instrument may be made into a robot-assisted tool by converting it into a miniature robotic system. While the robotic tool does other tasks, such as depth locking and tremor cancellation, the surgeon controls this instrument to perform a hands-on surgical procedure such as in "The Micron".

Popovic et al. [26] studied variations between FLACS (Femtosecond Laser-Assisted Cataract Surgery) and MCS (Manual Cataract Surgery) in the rate of posterior capsular opacification. Savastano and Rizzo [27] examined the viability and possible benefits of employing the Symani Surgical System, a revolutionary micro-surgical tele operated robot, in the area of ophthalmology. Qiu et al. [28] evaluated an iris tracker particularly for the ophthalmic robotic system. Iris tracking can help to locate the incision and safeguard patients. Robotic-assisted platforms are intended to improve surgical precision and shorten the learning curve. Ladha et al. [29] determined the robotic attributes contributed the most to these positional characteristics in both manual and aided activities. Maberley et al. [30] used Preceyes Surgical Robotic System (Eindhoven, Netherlands) will be contrasted against manually performed Internal limiting membrane (ILM) removing within the VRmagic Eyesi surgical simulation in Mannheim, Germany. To understand the potential advantages and constraints of developing technology, validation research are still needed for intraocular robotic surgery. Yang et al. [31] investigated master–slave system for robotics and safety control approaches with a simulated attachment and simulated feedback on force are provided to aid cataract surgery to increase CSR (Cataract Surgery Rate) and minimize the reliance of surgery on medical professional expertise. Robotic lasers can fix cylinder numbers to some extent with laser incisions and the aid of laser-assisted vision correction, an operation that would not be feasible using manual techniques.

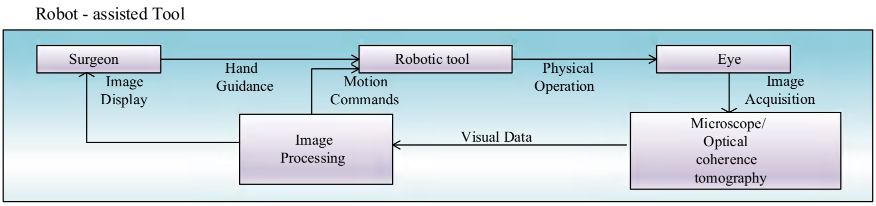


Fig. 2 The tactile navigation system's process flow in robot assisted tool [24]

Table 2 Analysis of intraocular surgery using robots

Researcher	Focus of research	Purpose
Edwards et al. [32]	Select a robot to inject recombinant tissue plasminogen	Inject recombinant tissue plasminogen for removing haemorrhage
Rahimy et al. [33]	Developed IRISS (Intraocular Robotic Interventional and Surgical System)	Perform both anterior and posterior segment surgery
Awad et al. [34]	Examined intraoperative factors that impact IO	Monitor IOP (Intraocular Pressure) fluctuations in patients
Bourcier et al. [35]	Examined the viability of simulated robot aided laser eye surgery	Perform experimental robot-assisted laser eye surgery
Nuzzi and Brusasco [36]	Examined the neurological surgery-areas of the brain that control visual acuity and ophthalmology	Analyse the merits and issues related of surgical in neurological surgery
Zhou et al. [37]	Do dual hand techniques in eye surgery using robotic ophthalmic endoscope holders	Prevent transferring, needle-like endoscopes from stepping into the retina
Hubschman et al. [38]	Tele-operation for access to ocular surgery in disadvantaged locations	Improve accuracy and dexterity, reduce time and minimize problems
Hubschman et al. [39]	Retina sensitivity	Provide necessary dexterity for sensitive intraocular procedures
Bourcier et al. [40]	Potential for employing the Da Vinci HD robotic surgical system for pterygium surgery	Outline and assess the use of a pig model for pterygium surgery

4.1 Intraocular

The current industry standard for robotic surgery in ophthalmology is da Vinci system. Few research articles explaining Robotic assisted surgery in intraocular was reviewed below (Table 2).

Majority of the articles evaluated above were related to the Robotic assisted surgery in intraocular. In the discipline of ophthalmology, da Vinci system is currently de facto standard robotic surgical system examined only one or two features and focused on image recognition of the area for surgery.

4.2 Cornea

Throughout a procedure, the strip of the cornea with the form of a disk is removed full-thickness corneal transplant, and it is then replaced with a donor piece of cornea of a same size. Sampat et al. [41] evaluated the most popular method of performing a radical prostatectomy (RP) is robotic-assisted laparoscopic radical prostatectomy. Tsirbas et al. [42] examined the application of an industrial surgical robot in ocular microsurgery. Antosh et al. [43] evaluated the prevalence of corneal abrasions following robotic/laparoscopic sacral colpopexies with that following vaginal apical suspensions, as well as to identify risk variables related to the occurrence of corneal abrasions. Pathirana and Kam [44] investigated the effects of anaesthesia as well as the particular difficulties and issues connected with these operations. Kwok et al. [45] examined degree of coordination between two independent operators, the effectiveness of surgical manipulation with and without CGC are assessed for managing a pair of bimanual surgical robots. Xiao et al. [46] investigated the causes of the challenges in regulating the depth of external device during corneal transplants from the viewpoint of the biomechanical characteristics of the cornea. Schweitzer and Kecik [47] examined phacoemulsification (PCS) as a standard care that improved over time and shown positive results in terms of anatomy, foresight, and index of refraction outcomes.

Majority of the articles evaluated above were related to the Robotic assisted surgery in cornea. A part of the patient's own cornea is replaced with donor corneal tissue during a cornea transplant. The translucent surface of the eye that resembles a dome is called the cornea through which the light penetrates the eye. It has a substantial impact on the eye's ability to see well.

4.3 Intravitrea

For the treatment of several retinal illnesses, such as diabetic retinopathy, age-related macular degeneration (AMD), and retinal vein occlusion, intravitreal medication administration has emerged as the gold standard. Since the invention of anti-vascular endothelial growth factor (VEGF) drugs, intravitreal injections have become much more common.

He et al. [48] demonstrated that the method may offer a means of attaining an intravitreal injection that is very autonomous and does not require surgery. Uy and Artiaga [49] evaluated the efficacy, operation time, and security results of two distinct intravitreal injections (IVI) procedures. Bourcier et al. [50] generated strabismus surgery using the new da Vinci Xi Surgical System, and the first instance of experimental eye muscle surgery using a surgical robot is being demonstrated. The efficacy of intravitreal ranibizumab (IVR) and intravitreal aflibercept (IVA) administered in accordance with the treat-and-extend (TAE) regimen on eyes with diabetic macular edema (DME) was studied by Chujo et al. [51]. Dar et al. [52] examined the clinical characteristics, evaluation of vision, and etiological agents of corneal inflammation following intravitreal injections (IVI) or cataract surgery. Starr et al. [53] evaluated intraoperative intravitreal anti-vascular endothelial growth factor (VEGF) injections serve as permanent treatment for patients with active diabetic macular edema (DME) undergoing cataract surgery. Nassiri et al. [54] evaluated the effectiveness of adding an intravitreal injection of triamcinolone acetonide-moxifloxacin (Tri-Moxi) to a regular eye drop practice in lowering postoperative inflammation, corneal edema, and the prevalence of high intraocular pressure (IOP) in cataract patients. Ren et al. [55] investigated reliability and utility of intravitreal conbercept, a synthetic fusion protein targeting vascular endothelial growth factors, after vitrectomy in the treatment of proliferative diabetic retinopathy without TRD. Gustavo Melo et al. [56] examined the information on syringe handling, anaesthesia, devices that help intravitreal injections, scleral piercing techniques utilising needles, silicone oil coating on syringes, safety concerns relating to materials and procedures, and needles, syringes, silicone oil coating, and sterilisation methods. When it comes to treating various retinal conditions, such as age-related macular degeneration (AMD) with neovascularization, diabetic retinopathy, and retinal vein occlusion, intravitreal medication administration has emerged as the gold standard.

4.4 Retina

Retina surgery, done on the retina of the eye is one of the most vital and critical which has to be done with high precision and accuracy. He et al. [57] introduced a master–slave robot system for retinal vascular bypass surgery. Yu et al. [58] demonstrated an experimental setup for a novel retinal micro-vascular surgery idea. Mango et al. [59] utilised digital ultrahigh resolution microscopes, overlaying intraoperative OCT data in real time, and automating laser assisted cataract surgery are just a few of the big advancements in our profession that have recently occurred. Roizenblatt et al. [60] most technically demanding minimally invasive surgical methods are vitreoretinal microsurgery. He et al. [61] analysed user behaviour in robot-assisted retinal surgery as approximated by three separate activities using four measurable parameters. Mitchell et al. [62] utilised a force sensor, the surgeon, the robot, and the steady-hand paradigm all share control of a tool that are coupled to the robot. Richa et al. [63] described a device that used the optical stereo microscope's visual input to identify accidental contact between surgical instruments and the retina. He et al. [64] illustrated how the force from the instrument to the tissue may be successfully controlled by the use of robotic assistance and force-to-audio sensory replacement. Horise et al. [65] introduced a "smart" light pipe that, by utilising an active, robot-assisted source of tailored illumination, enables genuine bimanual surgery.

Becker et al. [66] proposed an image-guided robotic system that enables scalable, smooth motion while reducing tremor. Gonenc et al. [67] integrated system that consists of a force-sensing motorised micro forceps and an active tremor-cancelling portable micromanipulator. Becker and Riviere [68] approached the individual-frame vessel detections to register and temporally fuse retinal vessel detections to map and localise retinal vessels. All the researches reviewed above requires the image of the retina be accurate for the surgical tool to perform the surgery.

5 Current Types of Surgical Robots Used in Ophthalmology and AI Usage

Da Vinci Surgical System's Steady-Hand Eye Robot, the Intraocular Robotic Interventional Surgical System (IRISS), smart tools, and Preceyes' B.V are solutions to overcome distance and a shortage of surgeons is cybersurgery.

Zhou et al. [69] evaluated the sub-retinal insertion robot system with intraoperative optical coherence tomography (OCT). Barthel et al. [70] analysed a hybrid parallel-serial micromanipulator that are conceived and developed for tele-manipulation employing a haptic device that uses feedback from force as its user interface. Sommer and Blumenthal [71] utilised tool for diagnosis and treatment of acute and chronic ocular disorders in the clinical context, integrating remote care into the daily workflow. Gharebaghi et al. [72] attempted to offer an evidence-based approach, albeit that may not be relevant to all centres and are not a confirmed medico-legal guideline. Singh et al. [73] regulated structures of a diameter of a few microns, where there was little tactile feeling, like the interior limiting membrane. Shi et al. [74] focused on the following areas as it examines the most current developments in various developing strategies for 3D form sensing in this field: Shape reconstruction techniques based on fibre optic sensors, electromagnetic tracking, and intraoperative imaging modalities. Ullrich et al. [75] focused on the application of robotic tools to facilitate eye surgery.

Tele ophthalmology is a branch of telemedicine that treats eyes using digital medical instruments and telecommunications technology. The use of tele ophthalmology is being made for remote learning, ocular disease detection, treatment, and surveillance, and also to provide patient access to eye doctors in remote places. In tele-operated robotic surgery, they may control by joystick uses microscope to display the image feedback.

Hickman et al. [76] analysed the prospective use of smartphone-shot surgical videos for supporting live tele-surgical consultation as well as self-recording and self-review for surgeon learning. Merzougui et al. [77] studied the presumption and concept to make this disease identification easier using tele-ophthalmology and cutting-edge telecommunications technologies. Chandra et al. [78] utilised innovation, interdisciplinary healthcare practitioners with diverse professional skills must operate efficiently to reduce patient contact time. Chen et al. [79] studied pressure from a wireless device attempting to find a paradigm for the single continuous intraocular pressure (IOP) monitoring of glaucoma patients. Laube et al. [80] utilized a tolerable surgical time frame, intraocular EPI RET3 implant is reproducible and viable. Rai et al. [81] studied that the construct validity for the EyeSI AR BIO simulator and suggests that it may be preferable to traditional BIO instruction for rookie ophthalmology trainees. Berger et al. [82] investigated the overlay of earlier taken photographic and angiographic pictures directly onto the contemporaneous slit-lamp bio-microscopic fundus image to guide therapy for macular disorders and to simplify real-time image measurement and comparison. Lai et al. [83] examined the existing restrictions on tele-neuro-ophthalmology use as well as its prospective application after the epidemic. The robotic surgery assistance provided by the daVinci Surgical

System enables surgeons to perform challenging minimally invasive surgical operations precisely and precisely. The apparatus is a state-of-the-art robotic platform designed to expand the surgeon's skills and provide them an open surgery option.

Cheng et al. [84] modified TernausNet-16 network, a surgical endoscope that was magnetically affixed underwent deep learning instrument tracking. Chen et al. [85] proposed spatio-temporal context learning and convolutional neural network (CNN) with a line segment detector (LSD) for visual tracking for two-dimensional tool identification.

Choi et al. [86] used CNNs to detect surgical instruments during laparoscopic surgery with YOLO architecture servers to track surgical tools in real time. Colleoni et al. [87] developed a 3D FCNN architecture to extract spatio-temporal information from laparoscopic videos for robotic MIS operations. Ni et al. [88] developed an attention-guided lightweight network named LWANet for real-time segmentation of surgical instrument. LWANet uses an encoder-decoder architecture with MobileNetV2 as the encoder and attention fusion block as the decoder. Deep learning models are challenging due to high computational costs and slow inference speed. Nakazawa et al. [89] suggested a convolutional neural network formed by bounding boxes for an actual time needle identification system. Shvets et al. [90] provided a deep learning robotic instrument segmentation solution based on the UNet network architecture, to address the binary segmentation problem. Zhang et al. [91] developed new GAN frameworks for image segmentation with unpaired data I2I. Unpaired image-to-image translation (I2I) techniques used to reduce the need for paired data for surgical tool segmentation. However, accurate surgical tool segmentation are challenging due to complex environment.

Zhao et al. [92] established Dense Convolutional Encoder–Decoder Network (DCED-Net) in an unsupervised deep learning network that extracts more discriminative characteristics from videos. Li et al. [93] suggested a deep CNN-based eye tracking method for controlling surgical robots. Laws et al. [94] used convolutional neural network and GoogLeNet to segment entire surgical views for marker less (CAOS) Computer Assisted Orthopedic Surgery systems. Marsden et al. [95] established Fluorescence Lifetime Imaging (FLIm) to diagnose oral and Oropharyngeal cancer surgery. Tayama et al. [96] studied autonomous placement of an eye surgery robot mounted micropipette are presented; particularly, the micropipette's shadow are employed for positioning in the depth direction. Using the visual serving technique, the micropipette was placed at the predetermined target point in the microscopic picture and then moved in the depth direction.

Jacob et al. [97] utilised a database of pictures of the seven common surgical tool types: hooks, forceps, hemostats, scissors, retractors, and scalpels. Zhao et al. [98] established the viability and promise of the suggested CNN-based approach to the learning of the surgeon's competence in ES and the perception of the unstructured surgical state. Kim et al. [99] used silicone eye phantom, an autonomous navigation system are tested in real-world tests and in simulation. Resnet-18 were used RMSE function for future-image prediction, and cross-entropy loss for baseline network training. Nguyen et al. [100] used to categorise voxels that contain different intensity information and are simple to recognise, whereas DNNs and CNNs (Convolution neural networks) are used to address voxels that are similar in appearance and typically are not recognised well enough by conventional methods.

The articles evaluated above were related to the related to the application of tele surgery in ophthalmology and Artificial Intelligence (AI) system in Robotics based ocular surgery. Massive neural networks with multiple layers of processing units are used in deep learning, which uses advancements in computer power and training techniques to find complex patterns in massive amounts of data. By using image recognition for tool detection and segmentation, surgical robots can be effectively assisted in carrying out more in precise surgery.

6 Performance Analysis

The performance analysed intricate and specific sub-stitch data, machine learning can accurately predict surgical experience mode choice is carried out using MAT Lab software. The dataset employed is STARE (Structured Analysis of the Retina) for the eye images. To accurately anticipate, machine learning and deep learning algorithms such as Gaussian Mixture Model (GMM), support vector machines (SVM), a custom-made convolutional neural network (CNN) is utilised. It operates by scanning the pathways of the eyes on the picture to obtain a series of eye projection locations.

The accuracy is analyzed and its graphical representation is illustrate in Fig. 3. Accuracy of HMM-SVM, GMM-SVM, CNN, Deep CNN and ResNet-18 is noted and that Deep CNN overcomes all other techniques, where it gives best results in classification, the accuracy was HMM-SVM 86%, GMM-SVM 90%, CNN 92%, Deep CNN 95% and ResNet-18 94%. Pattern recognition in images and videos is accomplished by deep convolutional neural networks (CNN or DCNN). DCNNs evolved from standard artificial neural networks and use a three-dimensional neural pattern modelled after the visual cortex. Performance between different model is compared using error metric and given in Fig. 4. For access mode the error value for proposed HMM-SVM and model such as GMM-SVM, CNN, Deep CNN and ResNet-18 is 0.1%, 0.08%, 0.05% and 0.06%. Then, in case of the error value attained for the proposed HMM-SVM is 0.14%. Through this comparison study, it is concluded that better performance is achieved using proposed regression model for mode choice analysis as error rate is minimal.

Performance of proposed model is compared and studied using precision value in Fig. 5. Precision value attained for the proposed Deep CNN is 94%.On the other hand precision value attained for some of the traditional regression model such as HMM-SVM, GMM-SVM, CNN, Deep-CNN and ResNet-18 mode is 82%, 89%, 90% and 92%. Based on this comparison study it is concluded that proposed model performs better in analysis by the visual cortex, have developed from typical artificial neural networks. Comparison study between the models using recall parameter is illustrated in Fig. 6. For access mode the recall value for proposed Deep CNN and traditional

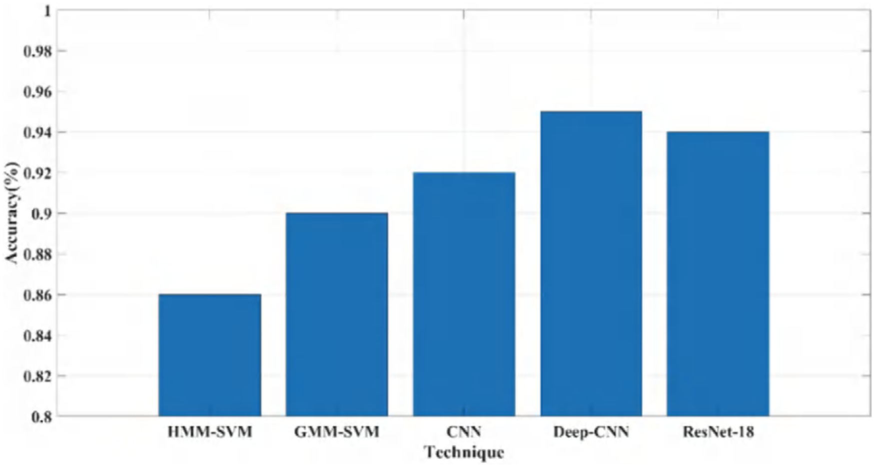


Fig.3 Accuracy comparison between SVM and CNN

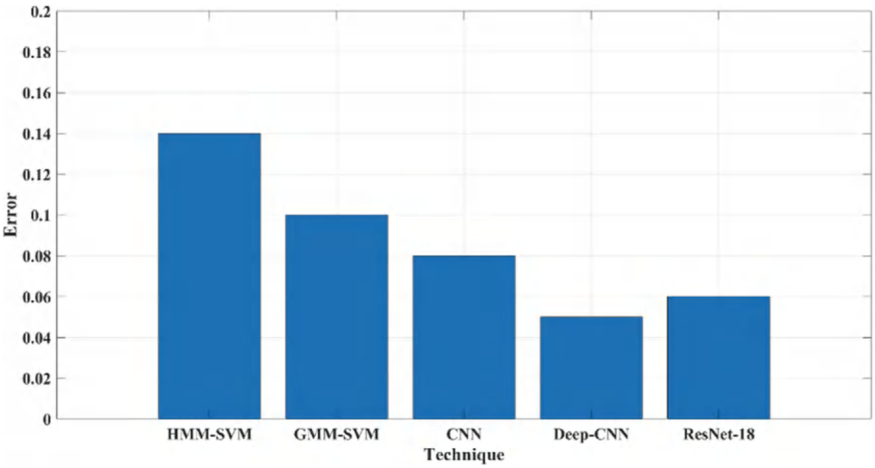


Fig.4 Error comparison between SVM and CNN

models such as HMM-SVM, GMM-SVM, CNN, and ResNet-18 are 84%, 87%, 91% and 92%. Through this recall comparison it is revealed that proposed Deep CNN regression model functions effectively in finding the visual cortex, have developed from typical artificial neural networks.

Figure 7 displays the comparison of specificity metric between model for both access and egress mode. Proposed Deep CNN 96% and traditional models such as HMM-SVM, GMM-SVM, CNN, and ResNet-18 are 84%, 88%, 92% and 95%. Through this specificity metric comparison it is revealed that proposed Deep CNN regression model functions is effective in finding the visual cortex, have developed

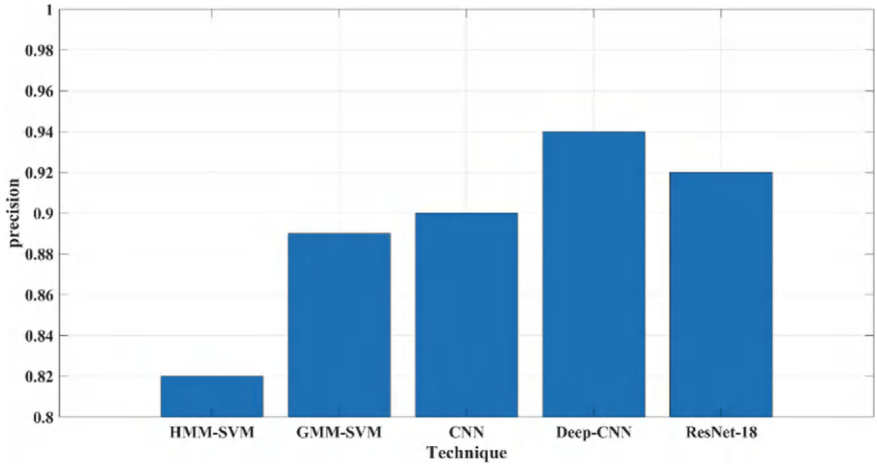


Fig. 5 Precision comparison between SVM and CNN

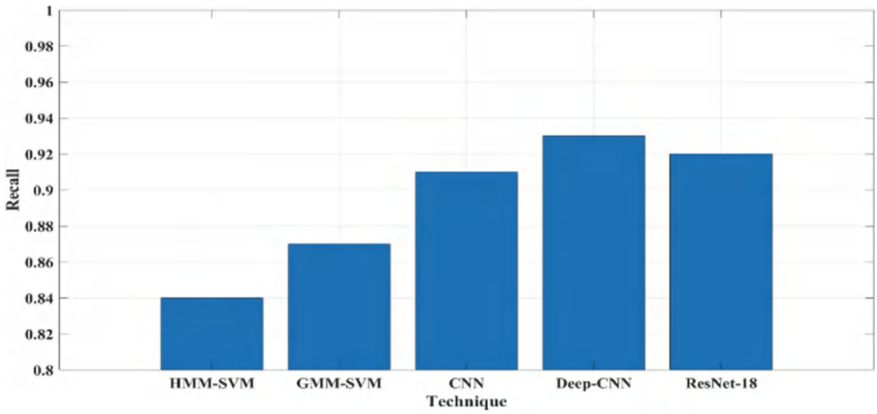


Fig.6 Recall comparison between SVM and CNN

from typical artificial neural networks. Figure 8 displays the comparison of execution time between the model for both access and egress mode. From the figure it is found that execution time estimated on using HMM-SVM is 25 s, calculated execution time for GMM-SVM, CNN, Deep CNN and ResNet-18 for access mode is 27s, 102s, 110s and 200 s. Based on this analysis it is concluded that time is lesser for HMM-SVM on comparison with other modes.

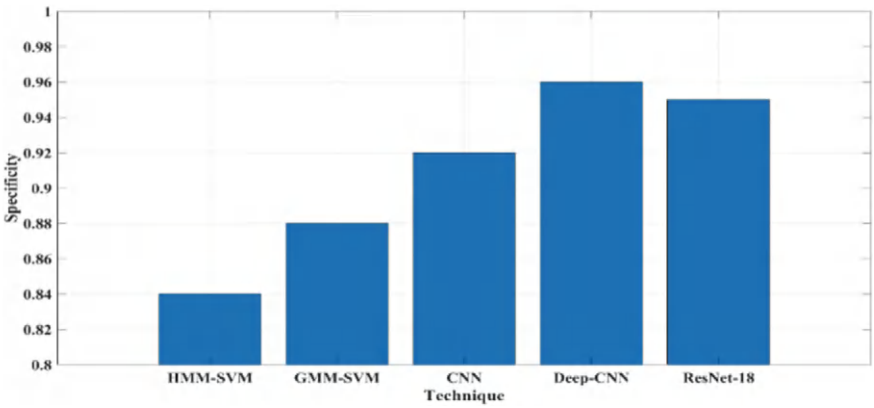


Fig.7 Specificity comparison between SVM and CNN

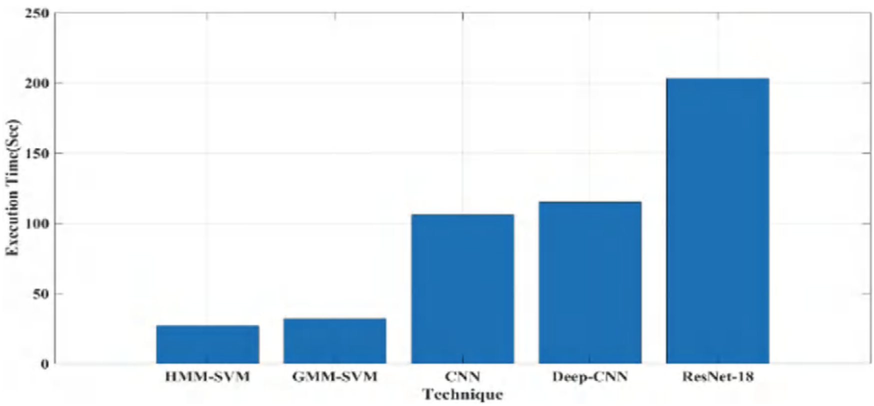


Fig.8 Execution time SVM and CNN

7 Conclusion

The intraocular convenience, instrument refining, and perception issues specific to ocular robotic surgery. Benefits of robotic surgery include better accuracy, improved flexibility in action, the removal of tremor, the capacity to navigate in a small anatomical region, decreased mistake, increased flexibility, and raised further physician risk. Deep learning and artificial intelligence have become key tools for capturing data from ocular surgery in order to analyse, instruct, and assist the surgeon in all aspects of surgical treatment. This study offered a technique for using CNN and SVM for identifying risk variables that are essential for early diagnosis to avoid the illness progressing to end-stage using deep learning algorithms. In automated ophthalmology analysis, deep learning AI algorithms displayed exceptional precision and good accuracy. AI-assisted examinations saw improvements in precision and accuracy of 95% and 94% in Deep CNN respectively. AI systems are extremely useful for illness detection, risk management, and image-based diagnosis. Due to the special characteristics of surgical perform, surgeons are in an ideal position to contribute to the next stage of AI, which aims to produce based on facts, immediate time clinical decision assistance to improve the treatment of patients and physician productivity. The paper only focusses on various robotic surgery techniques existing for ophthalmology and a study of the image recognition based on CNN and SVM is carried out on Stare dataset. The study could extend to various other features such as force detection, localization of the surgical tool, tremor detection and auto-correction etc. In addition, more datasets could be utilized and the research could be extended to other medical areas for different surgeries.

References

1. He, B., de Smet, M. D., Sodhi, M., Etminan, M., & Maberley, D. (2021). A review of robotic surgical training: establishing a curriculum and credentialing process in ophthalmology. *Eye*, *35*(12), 3192–3201.
2. Singh, S. K., Sharma, J., Joshua, L. M., Huda, F., Kumar, N., & Basu, S. (2022). Tele surgery and Robotics: Current Status and Future Perspectives. Telehealth/Telemedicine–The Far-Reaching Medicine for Everyone and Everywhere.
3. Peters, B. S., Armijo, P. R., Krause, C., Choudhury, S. A., & Oleynikov, D. (2018). Review of emerging surgical robotic technology. Surgical endoscopy, 32, 1636–1655.
4. Gamasu, R. (2015). Literature Review of Telemedicine System for Emergency Health Tribulations. Journal of Electronic and Electrical Engineering, 3(2), 163–170.
5. Ting, D. S. W., Pasquale, L. R., Peng, L., Campbell, J. P., Lee, A. Y., Raman, R., ... & Wong, T. Y. (2019). Artificial intelligence and deep learning in ophthalmology. British Journal of Ophthalmology, 103(2), 167–175.
6. Balyen, L., & Peto, T. (2019). Promising artificial intelligence-machine learning-deep learning algorithms in ophthalmology. The Asia-Pacific Journal of Ophthalmology, 8(3), 264–272.
7. Channa, R., Iordachita, I., & Handa, J. T. (2017). Robotic eye surgery. Retina (Philadelphia, Pa.), 37(7), 1220.
8. Christensen, H., Paz, D., Zhang, H., Meyer, D., Xiang, H., Han, Y., ... & Tang, S. (2021). Autonomous vehicles for micro-mobility. Autonomous Intelligent Systems, 1, 1–35.
9. Dawn, A. G., & Lee, P. P. (2004). Patient expectations for medical and surgical care: a review of the literature and applications to ophthalmology. Survey of ophthalmology, 49(5), 513–524.
10. Amirabdollahian, F., Livatino, S., Vahedi, B., Gudipati, R., Sheen, P., Gawrie-Mohan, S., & Vasdev, N. (2018). Prevalence of haptic feedback in robot-mediated surgery: a systematic review of literature. Journal of robotic surgery, 12, 11–25.
11. Alafaleq, M. (2023). Robotics and cybersurgery in ophthalmology: a current perspective. Journal of Robotic Surgery, 1–12.
12. Viscusi, E. R. (2008). Patient-controlled drug delivery for acute postoperative pain management: a review of current and emerging technologies. Regional Anesthesia & Pain Medicine, 33(2), 146–158.
13. Szold, A., Bergamaschi, R., Broeders, I., Dankelman, J., Forgione, A., Langø, T., ... & Vilallonga, R. (2015). European Association of Endoscopic Surgeons (EAES) consensus statement on the use of robotics in general surgery. Surgical endoscopy, 29, 253–288.

14. Lema, M. A., Laya, A., Mahmoodi, T., Cuevas, M., Sachs, J., Markendahl, J., & Dohler, M. (2017). Business case and technology analysis for 5G low latency applications. IEEE Access, 5, 5917–5935.
15. Camarillo, D. B., Krummel, T. M., & Salisbury Jr, J. K. (2004). Robotic technology in surgery: past, present, and future. The American Journal of Surgery, 188(4), 2–15.
16. Stoianovici, D. (2000). Robotic surgery. World journal of urology, 18, 289–295.
17. Hillel, A. T., Kapoor, A., Simaan, N., Taylor, R. H., & Flint, P. (2008). Applications of robotics for laryngeal surgery. Otolaryngologic Clinics of North America, 41(4), 781–791.
18. Hanly, E. J., & Talamini, M. A. (2004). Robotic abdominal surgery. The American journal of surgery, 188(4), 19–26.
19. Senapati, S., & Advincula, A. P. (2005). Telemedicine and robotics: paving the way to the globalization of surgery. International Journal of Gynecology & Obstetrics, 91(3), 210–216.
20. Thiel, D. D., & Winfield, H. N. (2008). Robotics in urology: past, present, and future. Journal of endourology, 22(4), 825–830.
21. Sim, H. G., Yip, S. K. H., & Cheng, C. W. S. (2006). Equipment and technology in surgical robotics. World journal of urology, 24, 128–135.
22. Xiong, L., Chng, C. B., Chui, C. K., Yu, P., & Li, Y. (2017). Shared control of a medical robot with haptic guidance. International journal of computer assisted radiology and surgery, 12, 137–147.
23. Lehman, A. C., Rentschler, M. E., Farritor, S. M., & Oleynikov, D. (2007). The current state of miniature in vivo laparoscopic robotics. Journal of robotic surgery, 1, 45–49.
24. Zemmar, A., Lozano, A. M., & Nelson, B. J. (2020). The rise of robots in surgical environments during COVID-19. Nature Machine Intelligence, 2(10), 566–572.
25. Camps, J. I. (2011). The use of robotics in pediatric surgery: my initial experience. Pediatric surgery international, 27, 991–996.
26. Popovic, M., Campos-M, X., Schlenker, M. B., & Ahmed, I. I. K. (2016). Efficacy and safety of femtosecond laser-assisted cataract surgery compared with manual cataract surgery: a meta-analysis of 14 567 eyes. Ophthalmology, 123(10), 2113–2126.
27. Savastano, A., & Rizzo, S. (2022). A Novel Microsurgical Robot: Preliminary Feasibility Test in Ophthalmic Field. Translational Vision Science & Technology, 11(8), 13–13.
28. Qiu, H., Li, Z., Yang, Y., Xin, C., & Bian, G. B. (2020). Real-time iris tracking using deep regression networks for robotic ophthalmic surgery. IEEE Access, 8, 50648–50658.
29. Ladha, R., Meenink, T., Smit, J., & de Smet, M. D. (2023). Advantages of robotic assistance over a manual approach in simulated subretinal injections and its relevance for gene therapy. Gene therapy, 30(3-4), 264–270.
30. Maberley, D. A., Beelen, M., Smit, J., Meenink, T., Naus, G., Wagner, C., & de Smet, M. D. (2020). A comparison of robotic and manual surgery for internal limiting membrane peeling. Graefe's Archive for Clinical and Experimental Ophthalmology, 258, 773–778.
31. Yang, Y., Jiang, Z., Yang, Y., Qi, X., Hu, Y., Du, J., ... & Liu, G. (2020). Safety control method of robot-assisted cataract surgery with virtual fixture and virtual force feedback. Journal of Intelligent & Robotic Systems, 97, 17–32.
32. Edwards, T. L., Xue, K., Meenink, H. C. M., Beelen, M. J., Naus, G. J. L., Simunovic, M. P., ... & MacLaren, R. E. (2018). First-in-human study of the safety and viability of intraocular robotic surgery. Nature biomedical engineering, 2(9), 649–656.
33. Rahimy, E., Wilson, J., Tsao, T. C., Schwartz, S., & Hubschman, J. P. (2013). Robot-assisted intraocular surgery: development of the IRISS and feasibility studies in an animal model. Eye, 27(8), 972–978.
34. Awad, H., Santilli, S., Ohr, M., Roth, A., Yan, W., Fernandez, S., ... & Patel, V. (2009). The effects of steep trendelenburg positioning on intraocular pressure during robotic radical prostatectomy. Anesthesia & Analgesia, 109(2), 473–478.
35. Bourcier, T., Chammas, J., Becmeur, P. H., Sauer, A., Gaucher, D., Liverneaux, P., ... & Mutter, D. (2017). Robot-assisted simulated cataract surgery. Journal of Cataract & Refractive Surgery, 43(4), 552–557.

36. Nuzzi, R., & Brusasco, L. (2018). State of the art of robotic surgery related to vision: brain and eye applications of newly available devices. Eye and brain, 13–24.
37. Zhou, D., Takeyama, H., Nakao, S., Sonoda, K. H., & Tadano, K. (2023). Real-time fundus reconstruction and intraocular mapping using an ophthalmic endoscope. The International Journal of Medical Robotics and Computer Assisted Surgery, e2496.
38. Hubschman, J. P., Son, J., Allen, B., Schwartz, S. D., & Bourges, J. L. (2011). Evaluation of the motion of surgical instruments during intraocular surgery. Eye, 25(7), 947–953.
39. Hubschman, J. P., Tsirbas, A. N. G. E. L. O., & Schwartz, S. D. (2008). Robotic surgery in ophthalmology. Retina Today, 81–84.
40. Bourcier, T., Nardin, M., Sauer, A., Gaucher, D., Speeg, C., Mutter, D. ... & Liverneaux, P. (2015). Robot-assisted pterygium surgery: feasibility study in a nonliving porcine model. Translational Vision Science & Technology, 4(1), 9–9.
41. Sampat, A., Parakati, I., Kunnavakkam, R., Glick, D. B., Lee, N. K., Tenney, M. ... & Roth, S. (2015). Corneal abrasion in hysterectomy and prostatectomy: role of laparoscopic and robotic assistance. Anesthesiology, 122(5), 994–1001.
42. Tsirbas, A., Mango, C., & Dutson, E. (2007). Robotic ocular surgery. British journal of ophthalmology, 91(1), 18–21.
43. Antosh, D. D., Whyte, T., Ezzell, A., Chen, B. A., Sokol, A. I., & Park, A. J. (2013). Incidence of corneal abrasions during pelvic reconstructive surgery. European Journal of Obstetrics & Gynecology and Reproductive Biology, 166(2), 226–228.
44. Pathirana, S., & Kam, P. C. A. (2018). Anaesthetic issues in robotic-assisted minimally invasive surgery. Anaesthesia and Intensive Care, 46(1), 25-35.
45. Kwok, K. W., Sun, L. W., Mylonas, G. P., James, D. R., Orihuela-Espina, F., & Yang, G. Z. (2012). Collaborative gaze channelling for improved cooperation during robotic assisted surgery. Annals of biomedical engineering, 40, 2156–2167.
46. Xiao, J., Chen, J., Li, M., & Zhang, L. (2023). Design and Experiment of an Ultrasound-Assisted Corneal Trephination System. Micromachines, 14(2), 438.
47. Schweitzer, C., & Kecik, M. (2023). Femtosecond laser-assisted cataract surgery: Update and perspectives. Frontiers in Medicine, 10, 290.
48. He, X., Luo, H., Feng, Y., Wu, X., & Diao, Y. (2022). An Uncalibrated Image-Based Visual Servo Strategy for Robust Navigation in Autonomous Intravitreal Injection. Electronics, 11(24), 4184.
49. Uy, H. S., & Artiaga, J. C. M. (2021). Comparison of Two Different Intravitreal Injection Techniques. Clinical Ophthalmology, 2383–2389.
50. Bourcier, T., Chammas, J., Gaucher, D., Liverneaux, P., Marescaux, J., Speeg-Schatz, C., ... & Sauer, A. (2019). Robot-assisted simulated strabismus surgery. Translational Vision Science & Technology, 8(3), 26–26.
51. Chujo, S., Sugimoto, M., Sasaki, T., Matsui, Y., Kato, K., Ichio, A. ... & Kondo, M. (2020). Comparison of 2-year outcomes between intravitreal ranibizumab and intravitreal aflibercept for diabetic macular edema with "treat-and-extend" regimen—its usefulness and problems. Journal of Clinical Medicine, 9(9), 2848.
52. Dar, N., Pillar, S., Friehmann, A., Belkin, A., & Ofir, S. (2020). Endophthalmitis after intravitreal injections versus cataract surgery: a 15-year cohort. International Ophthalmology, 40, 73–79.
53. Starr, M. R., Mahr, M. A., Smith, W. M., Iezzi, R., Barkmeier, A. J., & Bakri, S. J. (2021). Outcomes of Patients With Active Diabetic Macular Edema at the Time of Cataract Surgery Managed With Intravitreal Anti–Vascular Endothelial Growth Factor Injections. American Journal of Ophthalmology, 229, 194–199.
54. Nassiri, S., Hwang, F. S., Kim, J., LeClair, B., Yoon, E., Pham, M., & Rauser, M. E. (2019). Comparative analysis of intravitreal triamcinolone acetonide–moxifloxacin versus standard perioperative eyedrops in cataract surgery. Journal of Cataract & Refractive Surgery, 45(6), 760–765.
55. Ren, X., Bu, S., Zhang, X., Jiang, Y., Tan, L., Zhang, H., & Li, X. (2019). Safety and efficacy of intravitreal conbercept injection after vitrectomy for the treatment of proliferative diabetic retinopathy. Eye, 33(7), 1177–1183.

56. Melo, G. B., da Cruz, N. F. S., Emerson, G. G., Rezende, F. A., Meyer, C. H., Uchiyama, S., ... & Rodrigues, E. B. (2021). Critical analysis of techniques and materials used in devices, syringes, and needles used for intravitreal injections. Progress in Retinal and Eye Research, 80, 100862.
57. He, C. Y., Huang, L., Yang, Y., Liang, Q. F., & Li, Y. K. (2018). Research and realization of a master-slave robotic system for retinal vascular bypass surgery. *Chinese Journal of Mechanical Engineering*, *31*, 1–10.
58. Yu, H., Shen, J. H., Joos, K. M., & Simaan, N. (2013, May). Design, calibration and preliminary testing of a robotic telemanipulator for OCT guided retinal surgery. In *2013 IEEE International Conference on Robotics and Automation* (pp. 225–231). IEEE.
59. Mango, C. W., Tsirbas, A., & Hubschman, J. P. (2017). Robotic eye surgery. *Cutting edge of ophthalmic surgery: from refractive SMILE to robotic vitrectomy*, 145–152.
60. Roizenblatt, M., Edwards, T. L., & Gehlbach, P. L. (2018). Robot-assisted vitreoretinal surgery: current perspectives. *Robotic Surgery: Research and Reviews*, 1–11.
61. He, C., Ebrahimi, A., Roizenblatt, M., Patel, N., Yang, Y., Gehlbach, P. L., & Iordachita, I. (2018, August). User behavior evaluation in robot-assisted retinal surgery. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 174–179). IEEE.
62. Mitchell, B., Koo, J., Iordachita, I., Kazanzides, P., Kapoor, A., Handa, J. ... & Taylor, R. (2007, April). Development and application of a new steady-hand manipulator for retinal surgery. In *Proceedings 2007 IEEE International Conference on Robotics and Automation* (pp. 623–629). IEEE.
63. Richa, R., Balicki, M., Sznitman, R., Meisner, E., Taylor, R., & Hager, G. (2012). Vision-based proximity detection in retinal surgery. *IEEE transactions on biomedical engineering*, *59*(8), 2291–2301.
64. He, X., Gehlbach, P., Handa, J., Taylor, R., & Iordachita, I. (2014, August). Toward robotically assisted membrane peeling with 3-DOF distal force sensing in retinal microsurgery. In *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 6859–6863). IEEE.
65. Horise, Y., He, X., Gehlbach, P., Taylor, R., & Iordachita, I. (2015, August). FBG-based sensorized light pipe for robotic intraocular illumination facilitates bimanual retinal microsurgery. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 13–16). IEEE.
66. Becker, B. C., Voros, S., Lobes, L. A., Handa, J. T., Hager, G. D., & Riviere, C. N. (2010, August). Retinal vessel cannulation with an image-guided handheld robot. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* (pp. 5420–5423). IEEE.
67. Gonenc, B., Feldman, E., Gehlbach, P., Handa, J., Taylor, R. H., & Iordachita, I. (2014, May). Towards robot-assisted vitreoretinal surgery: Force-sensing micro-forceps integrated with a handheld micromanipulator. In *2014 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1399–1404). IEEE.
68. Becker, B. C., & Riviere, C. N. (2013, May). Real-time retinal vessel mapping and localization for intraocular surgery. In *2013 IEEE International Conference on Robotics and Automation* (pp. 5360–5365). IEEE.
69. Zhou, M., Yu, Q., Huang, K., Mahov, S., Eslami, A., Maier, M. ... & Nasseri, M. A. (2019). Towards robotic-assisted subretinal injection: A hybrid parallel–serial robot system design and preliminary evaluation. *IEEE Transactions on Industrial Electronics*, *67*(8), 6617–6628.
70. Barthel, A., Trematerra, D., Nasseri, M. A., Zapp, D., Lohmann, C. P., Knoll, A., & Maier, M. (2015, August). Haptic interface for robot-assisted ophthalmic surgery. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (pp. 4906–4909). IEEE.
71. Sommer, A. C., & Blumenthal, E. Z. (2020). Telemedicine in ophthalmology in view of the emerging COVID-19 outbreak. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *258*, 2341–2352.

72. Gharebaghi, R., Desuatels, J., Moshirfar, M., Parvizi, M., Daryabari, S. H., & Heidary, A. (2020). COVID-19: preliminary clinical guidelines for ophthalmology practices. *Gharebaghi R, Moshirfar M, Desuatels J, Parvizi M, Daryabari SH, Heidary F, COVID-19: Preliminary Clinical Guidelines for Ophthalmology Practices. Med Hypothesis Discov Innov Ophthalmol*, *9*(2), 149–158.
73. Singh, G., Jie, W. W. J., Sun, M. T., Casson, R., Selva, D., & Chan, W. (2022). Overcoming the impact of physiologic tremors in ophthalmology. *Graefe's Archive for Clinical and Experimental Ophthalmology*, *260*(12), 3723–3736.
74. Shi, C., Luo, X., Qi, P., Li, T., Song, S., Najdovski, Z. ... & Ren, H. (2016). Shape sensing techniques for continuum robots in minimally invasive surgery: A survey. *IEEE Transactions on Biomedical Engineering*, *64*(8), 1665–1678.
75. Ullrich, F., Fusco, S., Chatzipirpiridis, G., Pané, S., & Nelson, B. J. (2014). Recent progress in magnetically actuated microrobotics for ophthalmic therapies. *Eur. Ophthalmic Rev*, *8*, 120–126.
76. Hickman, M. S., Dean, W. H., Puri, L., Singh, S., Siegel, R., & Patel, D. (2022). Ophthalmic Tele surgery with a Low-Cost Smartphone Video System for Surgeon Self-Reflection and Remote Synchronous Consultation: A Qualitative and Quantitative Study. *Telemedicine Reports*, *3*(1), 30–37.
77. Merzougui, R., Hadjila, M., Benmessaoud, N., & Benaouali, M. (2018). Tele-Ophthalmology Android Application: Design and Implementation. *International Journal of Advanced Computer Science and Applications*, *9*(11).
78. Chandra, A., Romano, M. R., Ting, D. S., & Chao, D. L. (2021). Implementing the new normal in ophthalmology care beyond COVID-19. *European Journal of Ophthalmology*, *31*(2), 321–327.
79. Chen, P. J., Saati, S., Varma, R., Humayun, M. S., & Tai, Y. C. (2010). Wireless intraocular pressure sensing using microfabricated minimally invasive flexible-coiled LC sensor implant. Journal of Microelectromechanical Systems, 19(4), 721–734.
80. Laube, T., Brockmann, C., Roessler, G., Walter, P., Krueger, C., Goertz, M. ... & Bornfeld, N. (2012). Development of surgical techniques for implantation of a wireless intraocular epiretinal retina implant in Göttingen minipigs. Graefe's Archive for Clinical and Experimental Ophthalmology, 250, 51–59.
81. Rai, A. S., Rai, A. S., Mavrikakis, E., & Lam, W. C. (2017). Teaching binocular indirect ophthalmoscopy to novice residents using an augmented reality simulator. Canadian journal of ophthalmology, 52(5), 430–434.
82. Berger, J. W., & Shin, D. S. (1999). Computer-vision-enabled augmented reality fundus biomicroscopy. Ophthalmology, 106(10), 1935–1941.
83. Lai, K. E., & Ko, M. W. (2023). Can Tele-Neuro-Ophthalmology Be Useful Beyond the Pandemic? Current Neurology and Neuroscience Reports, 1–9.
84. Cheng, T., Li, W., Ng, W. Y., Huang, Y., Li, J., Ng, C. S. H., ... & Li, Z. (2021). Deep learning assisted robotic magnetic anchored and guided endoscope for real-time instrument tracking. IEEE Robotics and Automation Letters, 6(2), 3979–3986.
85. Chen, Z., Zhao, Z., & Cheng, X. (2017, October). Surgical instruments tracking based on deep learning with lines detection and spatio-temporal context. In 2017 Chinese Automation Congress (CAC) (pp. 2711–2714). IEEE.
86. Choi, B., Jo, K., Choi, S., & Choi, J. (2017, July). Surgical-tools detection based on Convolutional Neural Network in laparoscopic robot-assisted surgery. In 2017 39th annual international conference of the IEEE engineering in medicine and biology society (EMBC) (pp. 1756–1759). Ieee.
87. Colleoni, E., Moccia, S., Du, X., De Momi, E., & Stoyanov, D. (2019). Deep learning based robotic tool detection and articulation estimation with spatio-temporal layers. IEEE Robotics and Automation Letters, 4(3), 2714–2721.
88. Ni, Z. L., Bian, G. B., Hou, Z. G., Zhou, X. H., Xie, X. L., & Li, Z. (2020, May). Attention-guided lightweight network for real-time segmentation of robotic surgical instruments. In 2020 IEEE international conference on robotics and automation (ICRA) (pp. 9939–9945). IEEE.

89. Nakazawa, A., Harada, K., Mitsuishi, M., & Jannin, P. (2020). Real-time surgical needle detection using region-based convolutional neural networks. International Journal of Computer Assisted Radiology and Surgery, 15, 41–47.
90. Shvets, A. A., Rakhlin, A., Kalinin, A. A., & Iglovikov, V. I. (2018, December). Automatic instrument segmentation in robot-assisted surgery using deep learning. In 2018 17th IEEE international conference on machine learning and applications (ICMLA) (pp. 624–628). IEEE.
91. Zhang, Z., Rosa, B., & Nageotte, F. (2021). Surgical tool segmentation using generative adversarial networks with unpaired training data. IEEE Robotics and Automation Letters, 6(4), 6266–6273.
92. Zhao, H., Xie, J., Shao, Z., Qu, Y., Guan, Y., & Tan, J. (2018). A fast unsupervised approach for multi-modality surgical trajectory segmentation. IEEE Access, 6, 56411–56422.
93. Li, P., Hou, X., Wei, L., Song, G., & Duan, X. (2018, August). Efficient and low-cost deep-learning based gaze estimator for surgical robot control. In 2018 IEEE International Conference on Real-time Computing and Robotics (RCAR) (pp. 58–63). IEEE.
94. Laws, S. G., Souipas, S., Davies, B. L., & y Baena, F. R. (2020). Toward automated tissue classification for markerless orthopaedic robotic assistance. IEEE Transactions on Medical Robotics and Bionics, 2(4), 537–540.
95. Marsden, M., Weyers, B. W., Bec, J., Sun, T., Gandour-Edwards, R. F., Birkeland, A. C., ... & Marcu, L. (2020). Intraoperative margin assessment in oral and oropharyngeal cancer using label-free fluorescence lifetime imaging and machine learning. IEEE Transactions on Biomedical Engineering, 68(3), 857–868.
96. Tayama, T., Kurose, Y., Nitta, T., Harada, K., Someya, Y., Omata, S., ... & Mitsuishi, M. (2017). Image processing for autonomous positioning of eye surgery robot in micro-cannulation. Procedia CIRP, 65, 105–109.
97. Jacob, M. G., Li, Y. T., Akingba, G. A., & Wachs, J. P. (2013). Collaboration with a robotic scrub nurse. Communications of the ACM, 56(5), 68–75.
98. Zhao, Y., Guo, S., Wang, Y., Cui, J., Ma, Y., Zeng, Y. ... & Xiao, N. (2019). A CNN-based prototype method of unstructured surgical state perception and navigation for an endovascular surgery robot. Medical & Biological Engineering & Computing, 57, 1875–1887.
99. Kim, J. W., He, C., Urias, M., Gehlbach, P., Hager, G. D., Iordachita, I., & Kobilarov, M. (2020, May). Autonomously navigating a surgical tool inside the eye by learning from demonstration. In 2020 IEEE International Conference on Robotics and Automation (ICRA) (pp. 7351–7357). IEEE.
100. Nguyen, D. M., Vu, H. T., Ung, H. Q., & Nguyen, B. T. (2017, March). 3D-brain segmentation using deep neural network and Gaussian mixture model. In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 815–824). IEEE.

Examining Factors Influencing Behavioral Intentions to Use Artificial Intelligence in the Recruitment Processes of United Nations Agencies

Noor Al-Sabri, Redhwan Al-amri, and Gamal Alkawsi

Abstract This research examines the factors influencing the intention to use artificial intelligence (AI) technology in the recruitment process in UN agencies in Yemen, considering the influence of AI on the responsibilities of recruiters and human resources (HR) experts. Although AI technologies have made significant progress and can automate the recruitment process, limited research exists on recruiters' comprehension of AI systems and the factors driving AI adoption. This study aims to fill this gap in knowledge by examining the factors that influence the intention to use AI in the context of the recruitment process, integrating the Unified Theory of Acceptance and Use of Technology (UTAUT) and the Perceived Risk (PR) model as the underlying theories, providing an understanding of the factors affecting AI adoption in recruitment. Through a quantitative approach, data was collected from 93 HR and recruiting professionals in United Nations (UN) agencies in Yemen and analyzed using correlation and regression analyses. Results showed significant positive relationships between performance expectancy (PE), effort expectancy (EE), and the intention to use AI in recruitment, highlighting their roles in productivity and task completion. Social influence (SI) also positively impacts AI adoption. Conversely, time risk (TR) negatively affects the intention to use AI, underscoring the need to address and mitigate TR. No relationship was found between performance risk (PER) and behavioral intention (BI). These findings assist in mitigating time and performance concerns and implementing AI in recruitment with transparency.

N. Al-Sabri
Department of Business Administration, Lebanese International University, Sana'a, Yemen

R. Al-amri (✉)
Department of Applied Computing, Wales Institute of Science and Art, University of Wales Trinity Saint David, Swansea, UK
e-mail: r.al-amri@uwtsd.ac.uk

G. Alkawsi
Institute of Informatics and Computing in Energy (IICE), Universiti Tenaga Nasional, 43000 Kajang, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_20

1 Introduction

The recruitment process actively seeks out, locates, and employs people for particular positions or roles. It encompasses the whole recruitment process, from initiation to onboarding the recruited individual into the organization. The recruitment procedure takes time to complete [1]. Despite the development of AI technology, the conceptual framework of recruiting process approaches has remained constant, making the hiring process costly and time-consuming. Currently, the recruitment procedure in United Nations (UN) agencies in Yemen primarily involves human recruiters who manually sift through curriculum vitae, online profiles, and other sources to locate potential applicants. Recruiters assume responsibility for all preliminary communication, deliver feedback to unsuccessful candidates, and conduct interviews with prospective hires. Many recruitment and selection processes face difficulties because of the absence of AI techniques in those organizations. As mentioned by McRobert et al. (2018), managing all necessary tasks can be challenging and typically demands a significant amount of time and effort from each recruiter due to the limitations of human capabilities, such as biases, assumptions, and time constraints [2].

The HR industry has employed AI for multiple purposes, such as forecasting employee turnover, implementing chatbot systems for HR services, and validating applicants' credentials for resume screening. Initially, HR technology proved to be a proficient tool for evaluating applicant resumes and sourcing candidates. However, it has recently been acknowledged as an essential prerequisite for organizations seeking to grow and broaden their operations [3]. AI streamlines the recruitment process, such as screening resumes, shortlisting the fit candidates, scheduling interviews, and hiring suitable ones [4]. This is accomplished by automating a variety of procedures, including applicant matching and resume screening as well as the onboarding process [5]. AI supports fairness by evaluating all candidate attributes equally and excluding demographic information to avoid bias [6].

Despite the purported advantages of using AI technology in recruitment, there is a significant gap in understanding its adoption inside UN agencies, particularly in areas like Yemen. While prior studies have emphasized the benefits of AI in expediting HR processes and eliminating biases, little has been done to investigate the specific obstacles, issues, and factors impacting the inclination to use AI technology in the recruitment process. This research addresses a critical gap by investigating the factors involving in shaping the employees' opinion of using AI in UN agencies' recruitment processes and formulate a theoretical framework that defines the variables related to the users' inclination to utilize intelligence methods in the recruitment process within UN agencies, building upon performance and effort expectancy, as well as social influence factors from the UTAUT theory proposed by Venkatesh et al. [7]. Additionally, it investigates the anticipated risks associated with performance and the time risks from the perceived risk theory proposed by Featherman and Pavlou [8].

2 Related Studies

In China, Pan et al. (2022) explored the adoption of AI in employee recruitment and the impact of contextual factors. Their study contributes to the existing literature on HR management by introducing the Technology-Organization-Environment framework, validated through an online survey of 297 HR managers in Guangzhou [9]. Similarly, in Bangladesh, Alam et al. (2020) examined the inclination of HR employees to utilize AI in recruiting, employing the UTAUT model, and collecting data from various industry employees. Their findings highlight the positive effect of analyzed variables on the behavioral intention to use AI in recruitment, aiding HR professionals in understanding these influences [10]. In Korea, Lee et al. (2021) investigated AI-powered recruitment systems' growing prominence, aiming to address job market equity and impartiality. Using the Technology-Organization-Environment framework and TAM, they found that factors such as dependability, confidentiality, and emerging technology significantly impact the intention to use AI, with the firm's personnel count moderating these effects [11].

In Thailand, Wongras and Tanantong (2023) aimed to understand HR employees' views on using AI in recruitment, extending and customizing the UTAUT model. A survey of 364 HR professionals in Bangkok indicated that perceived value, autonomy, EE, and facilitating conditions (FC) influence AI adoption, while trust and social influence (SI) indirectly influence the decision [12]. In France, Horodyski, 2023 examined AI adoption in hiring, expanding the UTAUT model to include AI usage frequency and education. Their study, based on data from HR professionals in 15 countries, uncovered a robust and favorable relationship between PE and BI, but no significant association between EE and SI [13]. In Germany, Laurim et al. (2021) focused on AI acceptance criteria in hiring, using TAM theory to analyze individual and situational factors. Interviews with recruiters, managers, and applicants provided recommendations for leveraging AI in talent acquisition [14].

In the United Kingdom, Wright and Atkinson (2021) explored AI's influence on employers and candidates through an integrative literature review, including interviews and surveys. They identified themes such as risks and limitations, bias and inclusion, and technicalities and opportunities, highlighting AI's potential to streamline recruitment processes [15]. This collection of studies provides comprehensive insights into the various factors influencing AI adoption in recruitment across different cultural and organizational contexts, offering valuable recommendations for HR professionals and organizations worldwide.

Despite these comprehensive studies, there are significant gaps in the existing literature. Most research has been conducted in developed and developing countries, with no studies exploring the factors influencing AI adoption in recruitment in least developed countries like Yemen. Additionally, while many studies focus on the drivers of AI adoption, they often overlook the risk aspects that serve as barriers to technology adoption. Addressing these gaps is crucial for understanding the unique challenges and opportunities in AI adoption in diverse economic and cultural contexts, particularly in regions that have not been the focus of previous research.

3 Research Model and Hypothesis Development

This research focuses on the direct factors of UTAUT direct factors determining the desire to employ new technology, such as performance expectancy, effort expectancy, and social influence. Since the primary goal of this study is to scrutinize the constructs impacting users' behavioral intention to use AI in the recruitment process rather than the actual use of AI, the dependent variable (Use Behavior) and independent variable (facilitating condition) that have a direct relationship with (Use Behavior) were removed because they are outside the scope of this study. This theory of UTAUT was integrated with two constructs from perceived risk theory, including performance risk and time risk, according to the primary roles of recruiters as shown in Fig 1.

Table 1 shows the source of the selected variables and the reason behind the selection.

3.1 Research Hypothesis Development

Five elements from the UTAUT and PR models and five key research hypotheses are developed and examined to accomplish the objectives of the study, as listed below.

Performance Expectancy (PE)

PE is the degree to which a person thinks using the system will help them obtain significant rewards more effectively. Based on Venkatesh et al. (2003), PE is the most influential factor in determining one's intention to utilize new technologies. The foundation of this variable is based on perceived usefulness (TAM), external motivation (Motivation Model), compatibility with job requirements (Model of PC Utilisation), comparative advantage (Innovation Diffusion Theory), and expectations of outcomes (Social Cognition Theory) [7].

From the HR employees' viewpoint, this implies that the system helps them do duties more quickly, enhance their performance at work, raise productivity, increase

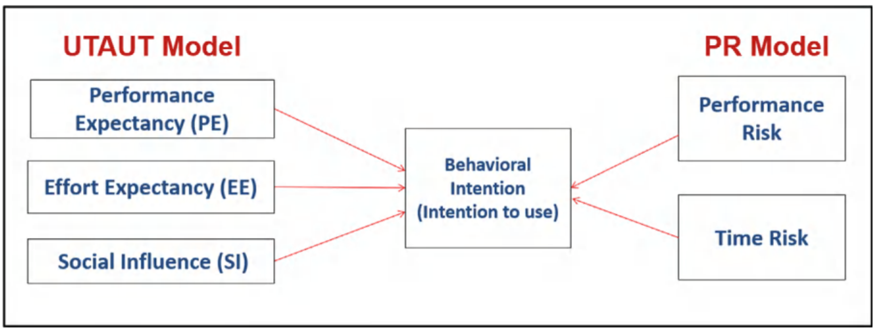


Fig. 1 Research model

Table 1 The selected constructs of the proposed model

Article	Model	Construct	Inclusion	Justification
Venkatesh et al. [7]	UTAUT	Performance expectancy	✓	One of the best indicators of intention is performance expectation, in which outcome expectations pertain to the consequences of action [7]
		Effort expectancy	✓	EE is viewed as a crucial factor that influences user towards the adoption of novel technological systems.; it means that the easier it is to use in individual expectations, the more intention to use [16]
		Social influence	✓	SI is a direct determinant of BI, in which the individual's behavior is influenced by how others will view them as a result of using the technology [7]
		Facilitating condition	×	FC is a direct determinant of use behavior, which is not in the scope of this study [13]
		Behavioral intention	✓	One's intention to take specific actions [13]
		Use Behavioral	×	Use behavior measures the actual usage of the new technology, which is out of the scope of this study [13]
Featherman and Pavlou [8]	Perceived risk theory	Performance risk	✓	Measure the technological PER since businesses need to be confident that the AI technology will be able to make accurate and fair decisions about who to hire [8]
		Psychological risk	×	This dimension is less likely to be a significant factor in using AI technology in recruitment
		Financial risk	×	While financial considerations are important for organizations implementing AI in recruitment, they may not directly impact individuals' inclination to utilize the technology

(continued)

Table 1 (continued)

Article	Model	Construct	Inclusion	Justification
		Privacy risk	×	Privacy risk is a significant consideration from the candidates' perspective regarding data protection and compliance; it holds less relevance within the scope of this research since it focuses on the organizational context rather than the consumer context
		Time risk	✓	TR addresses individuals' concerns about the potential time investment required to learn and use the technology and ensures that the processes that lead to the right results are performed effectively [8]
		Social risk	×	Since SI is already captured as a factor in UTAUT, it covers the social aspects of technology adoption, including the impact of social norms, subjective norms, and social support
		Overall risk	×	The overall risk is eliminated since it involve the other risks included above [17]

their effectiveness at work, and simplify work processes. From a performance viewpoint, a system is in use if it improves job effectiveness, frees up time for routine tasks, and enhances output quality and quantity for the same amount of effort. As stated by, Venkatesh, Morris, and Davis, 2003, performance expectancy is one of three direct factors in employing a new system in businesses. This supports the study's initial hypothesis that performance expectancy directly influences recruiters' intention to utilize AI tools for recruitment and selection [7].

Hypothesis 1 (H1): Performance Expectancy positively influences users' behavioural intention of using AI technology in recruiting.

Effort Expectancy (EE)

In the UTAUT model, the EE is considered another essential factor influencing users' intentions to adopt new technologies. It is defined as the level of ease associated with using the system; it was discovered that the influence of EE on behavioral intention varied across gender and age, with the effect being most significant for older women who are just beginning to use the system [7]. This factor was derived from the TAM aspect of perceived ease of use. According to Davis et al. (1989), a user-friendly application is more likely to be accepted by users [18]. According to a related study by Davis et al. (1989), effort-oriented constructs are predicted to be more prominent at the beginning of a new behavior when process difficulties arise [18]. Recruiters perceive the system as easy to learn, use, and navigate, providing flexibility and user-friendliness. Thus, the following hypothesis was suggested.

Hypothesis 2 (H2): Effort Expectancy positively influences users' behavioural intention of using AI technology in recruiting.

Social Influence (SI)

SI is when a person's actions are affected by how they think others will view them as a result of using the technology, as stated by Venkatesh et al. [7]. It is comparable to the "subjective norm" concept described in TAM 2, an extension of TAM. According to Moore and Benbasat,1996, the term "image" pertains to how an individual's social standing or reputation is perceived to be enhanced by adopting a technological breakthrough. Subjective norms and images, also known by other terms, encompass the notion that an individual's conduct is influenced by their perception of how others would see them when they use technology [19]. In TAM 2, the subjective norm directly impacts utilization intentions, perceived usefulness, and perceived ease of use of systems. However, none of the Social Influence constructs are essential when participation is voluntary. According to Venkatesh et al. (2003), subjective norms significantly influence perceived usefulness through internalization and identification, particularly in the early stages of experience, enhancing job performance [7].

Although there is abundant literature on different facets of SI on individual behavior, there is limited knowledge regarding the impact of adopting AI on HR employees and their inclination to utilize and integrate this technology in the recruitment process [13]. Nevertheless, Alam et al. (2020) mentioned that solid evidence suggests that social influence strongly predicts adopting technology such as AI [10]. Therefore, based on the discussions, the following hypothesis has been formulated.

Hypothesis 3 (H3): Social Influence positively influences users' behavioural intention of using AI technology in recruiting.

Perceived Risk (PR)

Featherman and Pavlou (2003) found that the PR theory is the ambiguity over the potentially unfavorable outcomes of utilizing a new product or service [8]. Raymond Bauer introduced the Perceived Risk Theory in 1960 to explain customer behavior and the factors influencing their decision-making process [20]. Consumer researchers most frequently utilize the idea of Perceived Risk, which defines risk in terms of how consumers perceive the unfavorable effects of purchasing a good or utility. Consumer researchers implicitly presume that each purchase event or use of a new service is uncertain regarding both likelihood and consequence [21]. Featherman and Pavlou (2003) mentioned that PR is typified into seven distinct aspects: PR, Psychological Risk, Financial Risk, Privacy Risk, TR, Social Risk, and Overall Risk [8]. Numerous studies on consumer behavior support the use of these risk factors to understand customer evaluations and purchases of goods and services. Two variables, performance and time were chosen for this study as discussed below:

Performance Risk: This is a crucial dimension to include in the research, as it directly relates to individuals' concerns about AI technology's ability to perform as expected in the recruitment process [22]. As AI technology plays a significant role in automating and augmenting recruitment tasks, individuals may have concerns about its accuracy, reliability, and effectiveness. Examining Performance Risk enables gaining insights into how individuals' perceptions of the technology's performance influence their intention to use it in recruitment.

Time Risk: is highly relevant in the context of AI technology in recruiting. Concerns about losing time learning the new technology and the potential mistakes in selecting the right person for a job are essential considerations for individuals when they intend to use AI technology [22]. Time issues arise due to the unfavorable outcomes that lead employees to perform the recruitment process manually, in addition to the time required to learn and adapt the new technology to the organization's needs.

Hypothesis 4 (H4): Performance Risk negatively influences users' behavioural intention of using AI technology in recruiting.

Hypothesis 5 (H5): Time Risk negatively influences users' behavioural intention of using AI technology in recruiting.

4 Methodology

The study employed a quantitative research methodology. Individual HR is the unit of analysis that examines employees' attitudes and beliefs. The primary theories used in the research are UTAUT and Perceived Risk, along with their respective predictors: PE, EE, and SI, as well as PER and TR. These factors were measured using a web-based. The survey commenced with a concise preamble elucidating the background of the questionnaires and the objective of the study. Subsequently, the participants proceeded to the questionnaire, comprising 21 questions specifically formulated to gather data on the inclination to utilize artificial intelligence in the realm of recruiting. The first part (Questions 1 to 4) covered the user's intention and performance expectancy of using AI in recruitment. The second part (questions 5 to 8) related to the effort expected by the HR employees using AI in the recruitment process. The third part (questions 9 to 11) addresses the participants' perceptions in terms of social influence and how AI will view them among their friends. The fourth part (questions 12 to 15) showed how much the participants intend to use AI in recruitment in the future. These four parts addressed the participants' perceptions of using AI in recruitment in terms of the factors of UTAUT. After that, the questions addressed participants' perceptions regarding the factors of PR. The fifth part (Questions 16 to 18) related to the performance risk expected by the participants. The sixth part (Questions 19 to 21) measures how participants anticipated the time risk of using AI in recruitment. Finally, questions 22 to 26 covered the demographic of the respondents in terms of gender, age, education, years of experience, and HR role. The participant responses were measured using 5-point Likert scale questions, with response options ranging from 1 (strongly disagree) to 5 (strongly agree).

The population is defined as a group of UN agencies' HR employees using traditional recruitment systems who are not yet AI technology users. The population is almost 100 employees, as the UN in Yemen states that 14 UN agencies operate in Yemen. The sampling table developed by Krejcie and Morgan simplifies determining the appropriate sample size by offering a reliable decision model [23]. The researchers found that the UN Development Program (UNDP) has 6 HR employees, and based on this fact, it was estimated that each HR department in UN agencies has roughly 6 HR employees. Therefore, the appropriate sample size was 80, following the guidelines outlined in the sample size determination using Krejcie and Morgan's Table.

Non-probability sampling was chosen for the study because it involves a non-random sample selection. A non-probability sampling technique, purposive sampling, was adopted according to the researcher's subjective judgment of who has the characteristics, knowledge, or experience associated with the research study. In addition, the researcher utilized a snowball sampling technique by asking the participants to recommend other HR employees from different UN agencies who might be interested in participating in the study. A pilot study is conducted to ensure the usefulness of the questionnaire. The criterion for inclusion in this study is being an HR employee in UN agencies.

Statistical Package for the Social Sciences (SPSS) was used to analyze the response and evaluate the proposed model, Subsequently, the data was subjected to analysis utilizing the Structural Equation Modeling (SEM) approach using the SmartPLS 4 software [24]. SPSS provides a display of the mean, standard deviation (SD), and other statistical measures of the data. Correlation analysis was used to determine the direction and strength of the constructs of UTAUT and PR. Regression analysis was used to examine the relationship between the independent and dependent variables. [25, 39, 40].

5 Findings and Analysis

5.1 Descriptive Analysis

This section covers a descriptive analysis of the respondents' profiles.

Demographic Profile of the Respondents

The descriptive analysis of the participant profiles is covered in this subsection. To yield significant insights, the data from the participants is compiled using fundamental descriptive statistics such as frequencies and percentages. Table 2 exhibits the variation of demographic profiles, including frequencies and percentages, among the study population participants.

More male participants than female participants have answered the survey, as indicated by the data in the table, which reveals that 50 (53.8 percent) of the participants are male and the remaining 43 (46.2 percent) are female. Regarding the sample's age distribution, 54.8 percent are between the ages of 18 and 35, 38.7 percent are between the ages of 36 and 45, 5.4 percent are between the ages of 46 and 65, and 1.1 percent are between 18 and 25 years. In terms of the level of education, 1 participant (1.1 percent) has a PhD, whereas 42 participants (45.2 percent) hold a master's degree. Meanwhile, 50 participants (53.8%) have a bachelor's degree. In terms of years of experience, Table 2 shows that two individuals (2.2%) have from 1 to 2 years of experience, while 23 individuals (24.7%) have 3 to 6 years of experience. 33 individuals (35.5%) have experience of 7 to 9 years, and those who have more than 10 years

Table 2 Demographic profile

	Criteria	Frequency	Percent
Gender	Male	50	53.8
	Female	43	46.2
	Total	**93**	**100.0**
Age	18 to 25 years old	1	1.1
	26 to 35 years old	51	54.8
	36 to 45 years old	36	38.7
	46 to 65 years old	5	5.4
	Total	**93**	**100.0**
Education	Bachelor's degree	50	53.8
	Master's degree	42	45.2
	Ph.D	1	1.1
	Total	**93**	**100.0**
Experience	1 to 2 years	2	2.2
	3 to 6 years	23	24.7
	7 to 9 years	33	35.5
	More than ten years	35	37.6
	Total	**93**	**100.0**
Position	Compensation and benefits	4	4.3
	Employee relations	3	3.2
	HR general	38	40.9
	HR manager	16	17.2
	Talent acquisition/recruitment	30	32.3
	Training and development	2	2.2
	Total	**93**	**100.0**

of experience are 35 people (37.6%). HR roles are distributed across six different categories. The most frequent category is "Talent Acquisition/Recruitment," with 34 entries (36.6%). It is followed by "HR General" with 34 entries (36.6%) and "HR Manager" with 16 entries (17.2%). The least frequent categories are "Employee Relations," with three entries (3.2%), and "Training and Development," with two entries (2.2%).

5.2 *Measurement Model*

Confirmatory Factor Analysis

Confirmatory factor analysis (CFA) is essential for investigating the compatibility between a construct's measurement items and the intended interpretation of those items within a study. Specifically, The hypothesized structure of the proposed theoretical model is used to determine the congruence between empirical observations and theoretical expectations. CFA is employed to achieve this. Additionally, it functions as an instrument for assessing the congruence between the theoretical constructs and empirical observations, consequently, it provides evidence of the construct validity and reliability of the research's measurement tools.

Reliability Testing

Reliability means that when a study is repeatable, it should produce the same outcomes if carried out again [26]. This test is done using Cronbach's alpha which involves determining the Cronbach's alpha coefficients for each individual factor. The Cronbach's alpha values of 0.871, 0.847, 0.815, 0.861, and 0.803 for PE, EE, BI, PER, and TR, respectively, indicate satisfactory internal consistency. However, regarding SI, the Cronbach's alpha coefficient is 0.660. This value indicates moderate internal consistency, suggesting that there may be a few issues with the scale's reliability. The researcher sees that the items included in the SI construct "I find AI trendy" and "The use of AI gives me a professional status" capture individuals' perceptions and attitudes towards AI. These two items may introduce some response variability due to the subjective nature of opinions on trends and professional status. Also, the item "People who influence my behavior think that I should use AI" may be influenced by individual differences in the social networks and opinions of those surrounding the respondents. These items, influenced by different aspects of SI, exhibit more significant variability than other variables, such as PE or EE, which may contribute to the slightly lower internal consistency (Table 3).

Convergent Validity: Average Variance Extracted

The term "validity" refers to how well the research's arguments, interpretations, and findings support the topic to which they are supposed to be related.

Convergent validity: It evaluates the degree of convergence among items that assess a common construct. Researchers employ the average variance extracted as

Table 3 Cronbach's alpha test

Variable	Cronbach's alpha	Number of items
PE	0.871	4
EE	0.847	4
SI	0.660	3
BI	0.815	4
PFR	0.861	3
TR	0.803	3

a criterion to establish convergent validity [27]. The average variance extracted is a standard metric for defining convergent validity at the construct level [27]. Discriminant validity refers to correlations between different constructs or factors. The goal is to determine whether the measures designed to assess other constructs are less correlated than those that assess the same construct.

Table 4 displays a correlation matrix representing the interrelationships among PE, EE, SI, BI, PER, and TR. As indicated by the figures, the average correlations between the items that comprise each construct are 0.728, 0.759, 0.796, 0.836, 0.627, and 0.755, respectively; these values represent the square root of the average variance extracted on the diagonals. The findings of the moderate correlation indicate strong convergent validity and a significant positive correlation among the items within each construct, implying that the convergent validity of the conceptual model construct is satisfied.

Discriminant validity is achieved when the average correlation between each construct and the relevant items (shown by the diagonal bold values) is more significant than the correlation between the constructs and other constructs (represented by the off-diagonal values). It can be observed from Table 4 that every value in the table that is diagonal is more significant than every value that is not diagonal. The table generally shows that the constructs have a high degree of discriminant validity, which results in the achievement of the discriminant validity of every construct.

Table 4 Convergent and discriminant validity

	BI	EE	PE	PFR	SI	TR
BI	**0.728**					
EE	0.549	**0.759**				
PE	0.768	0.337	**0.796**			
PFR	−0.364	−0.138	−0.368	**0.836**		
SI	0.826	0.519	0.850	−0.253	**0.627**	
TR	−0.363	−0.162	−0.204	0.640	−0.107	**0.755**

5.3 Regression Analysis

Regression analysis is a statistical methodology utilized to scrutinize the interdependence between a dependent variable and independent variables to comprehend how alterations in these variables impact the dependent variable. However, assessing the validity of several underlying assumptions is essential before conducting multiple regression analyses [28]. These assumptions, like linearity, independence of observations, homoscedasticity, and normality of residuals, ensure that the results and interpretations derived from the analysis are reliable [29].

Linearity

It is assumed that a linear relationship exists between the independent variables and the dependent variable. This linearity assumption ensures unbiased estimates and valid inferences about the population. When the linearity assumption is violated, it can lead to biased coefficient estimates and inaccurate predictions.

The analysis examined the relationships between the BI variable and several independent variables. Table 5 shows no significant deviations from linearity were observed for any variables since their associated p-values are more significant than the conventional significance threshold, typically set at 0.05. All the results support the presence of linear relationships between these variables. Based on these findings, it can be concluded that the observed relationships in the sample data are likely to represent the actual relationships in the study population.

Normal Distribution of Residuals

A normal probability plot of residuals is a common way to ensure whether the errors in a linear regression model follow a normal distribution. When plotting the values of the residuals in a regression analysis, a distribution that approximately resembles a bell-shaped curve is expected to be observed, as errors tend to follow a normal distribution when they occur. [30]. Figure 2 shows that the presence of a normal distribution in the residuals signifies that the model aligns well with the data and that the errors exhibit both randomness and independence. In simpler terms, the normal distribution of residuals indicates that the model is a suitable match for the data. It also implies that the errors within the model occur randomly and independently. As a result, the statistical tests conducted on the model produce reliable and accurate results.

Table 5 Regression analysis

Path	F	P
BI * PE	0.337	0.968
BI * EE	1.265	0.264
BI * SI	0.974	0.456
BI * PER	1.050	0.406
BI * TR	1.745	0.092

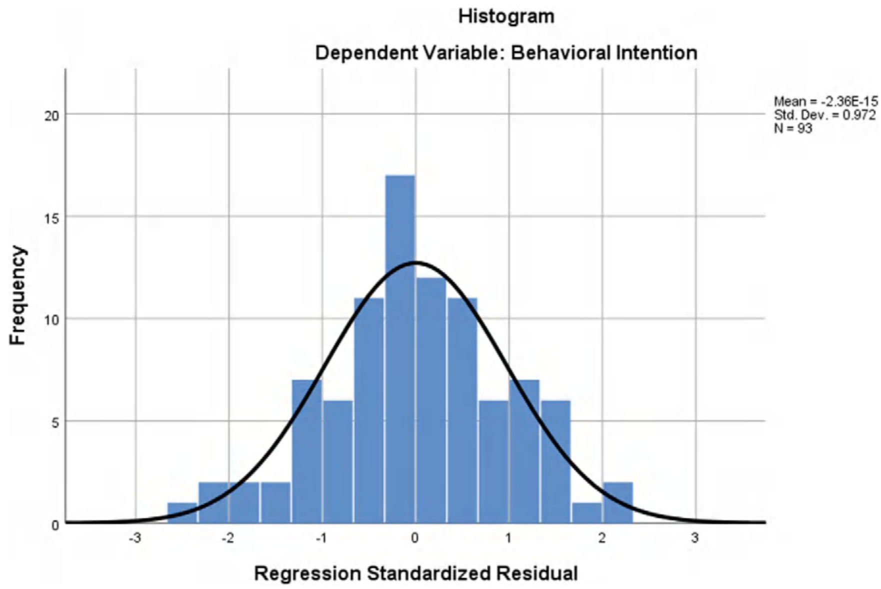


Fig. 2 Regression standardized residual

Homoscedasticity

Homoscedasticity pertains to the state where the dispersion of errors remains uniform or consistent across all levels of the independent variable. On the other hand, heteroscedasticity occurs when the variability of errors differs among different values of the independent variable. While reasonable heteroscedasticity has a minimal impact on the significance tests, significant heteroscedasticity can substantially skew research findings and weaken the analysis [29].

Figure 3 demonstrates that the distribution of errors above and below the regression line is symmetrical, and there is no correlation between the standardized predictive values of the regression and the regression residuals. This suggests that the assumption of homoscedasticity is met, and this indicates that the variability of the errors is constant across different levels of the predictors, supporting the assumption of homoscedasticity.

Independence of Errors

The independence of errors is assumed in the context of regression analysis, which means that the errors (residuals) in the model are not correlated with each other [31]. Violating this assumption results in biased estimates of standard errors, while the estimates of the regression coefficients remain unbiased but less efficient [31].

The Durbin-Watson test is a statistical technique to detect autocorrelation in regression analysis residuals. The Durbin-Watson statistic exhibits a numerical range from 0 to 4; a value close to 2 signifies the absence of autocorrelation. The Durbin-Watson statistic of 1.72 in Table 6 is near 2, suggesting that the residuals do not show

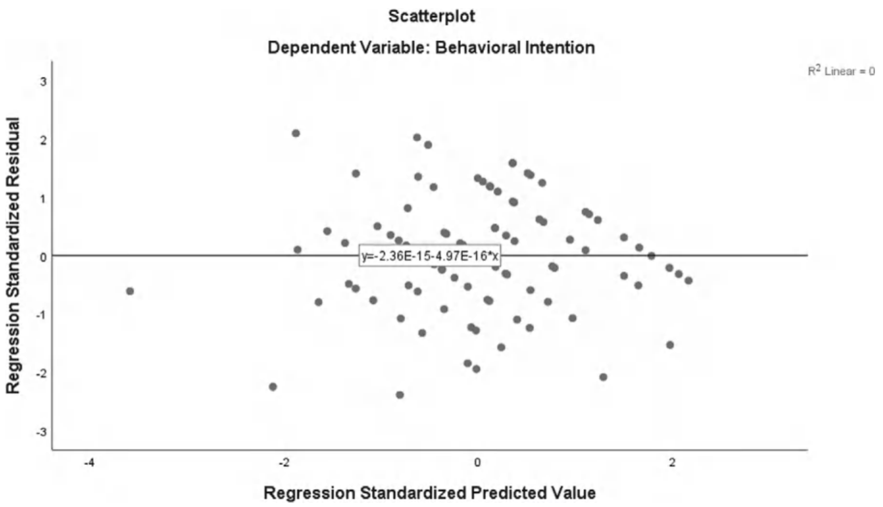


Fig. 3 Regression standardized predicted value

Table 6 Independence of errors

Model	Durbin-Watson
1	1.723

any signs of autocorrelation. The absence of autocorrelation is a desirable finding for the analysis because residual autocorrelation can affect the validity of regression results; hence, the fourth assumption was met.

Multicollinearity

When there are correlations between the predictors, the relation between two predictor variables is called collinearity, and when there are correlations between more than two predictors, it is called multicollinearity. Collinearity statistics are essential when evaluating multicollinearity among independent variables in a regression study.

PE, EE, SI, PR, and TR are the five independent variables whose variance inflation factor and tolerance values are displayed in Table 7. Various statistics are crucial markers of the correlation between multiple variables. The variance inflation factor values greater than 3 indicate some degree of multicollinearity, while values less than 3 imply no multicollinearity. All the variance inflation factor values in this situation are fewer than three and relatively close to 1, indicating no multicollinearity among the independent variables. This result is consistent with the tolerance values, which all exceed 0.20. These findings verify that every variable affects the dependent variable independently of the other factors.

Table 7 Collinearity test

Model	Collinearity statistics	
	Tolerance	Variance inflation factor
PE	0.553	1.807
EE	0.858	1.166
SI	0.559	1.787
PER	0.632	1.583
TR	0.675	1.481

a. Dependent Variable: BI

Model Quality Test

The R^2 value indicates the amount of variation in the dependent variables accounted for by the independent variables. A higher R^2 value signifies improved predictive ability of the structural model. It is essential to have sufficiently large R^2 values for meaningful predictive power [32]. According to Falk and Miller (1992), a construct's explained variance can only be considered sufficient if its R^2 values are more than or equal to 0.10 [33].

Table 8 shows that the model with an R^2 value of 0.567 can explain 56.7% of the variation in AI intention to use. The adjusted R^2, which considers the number of independent variables, is also 0.542, providing a more accurate estimation of the model's predictive ability; this indicates that a substantial portion of the variability in AI intention to use can be attributed to the independent variables. These findings suggest that the model has sufficient power and accuracy to elucidate the correlation between the IV and DVs.

Model Fit

Model fit is an aspect of regression analysis as it identifies the reliability and validity of the model's predictions and inferences. A good model fit indicates that the selected independent variables adequately clarify the variation in the dependent variable [34].

The ANOVA Table 9 provides information about the model fit for the regression analysis conducted on the dependent variable, "BI", with the predictors being "TR", "SI", "EE", "PER", and "PE". Based on the ANOVA results, the model shows a good fit in explaining the "BI" variable, as evidenced by the significant F value and the substantial proportion of variability (F-statistic = 22.776, $p < 0.001$).

Table 8 Model summary

Model	R	R square	Adjusted R square	Std. Error of the estimate
1	0.753[a]	0.567	0.542	0.37995

a. Predictors: (Constant), TR, SI, EE, PER, PE

Table 9 Anova test

ANOVA						
Model		Sum of squares	df	Mean square	F	Sig.
1	Regression	16.440	5	3.288	22.776	0.000[b]
	Residual	12.560	87	0.144		
	Total	29.000	92			

a. Dependent Variable: BI
b. Predictors: (Constant), TR, SI, EE, PER, PE

Model Coefficients

As emphasized by Thompson and Borrello (1985), the importance of structure coefficients in regression analysis, particularly when predictor variables are correlated. These coefficients can help interpret the regression results [35]. As shown in Table 10 and Fig. 4, the analysis revealed that Performance Expectancy (PE) has a significant positive relationship with Behavioral Intention (BI), with a beta value of 0.403 ($t = 4.253$, $p < 0.001$). This suggests that PE accounts for 40.3% of the variation in BI, indicating that a higher level of PE regarding the use of AI in recruitment processes leads to a stronger intention to adopt AI. Similarly, Effort Expectancy (EE) was found to positively influence BI, with a beta value of 0.225 ($t = 2.960$, $p = 0.004$), explaining 22.5% of BI. This significant relationship implies that the perceived ease of using AI in recruitment is associated with a stronger intention to use AI, thereby supporting the hypothesis.

Social Influence (SI) also shows a significant positive relationship with BI, with a beta value of 0.248 ($t = 2.627$, $p = 0.010$), indicating that SI accounts for 24.8% of the variation in BI. This suggests that social factors, such as peer influence and societal norms, play a crucial role in shaping the intention to use AI in recruitment, thus supporting the hypothesis. In contrast, Performance Risk (PER) does not show a significant relationship with BI, as evidenced by a beta value of −0.004 ($t = -0.046$, $p = 0.964$). This lack of significance indicates that PER does not contribute

Table 10 Coefficient analysis

Model	Unstandardized coefficients		Standardized coefficients	t	Sig.
	B	Std. Error	Beta		
PE	0.307	0.072	0.403	0.253	0.000
EE	0.204	0.069	0.225	0.960	0.004
SI	0.232	0.088	0.248	2.627	0.010
PER	0.003	0.060	−0.004	−0.046	0.964
TR	0.122	0.058	−0.181	−2.109	0.038

a. Dependent Variable: BI

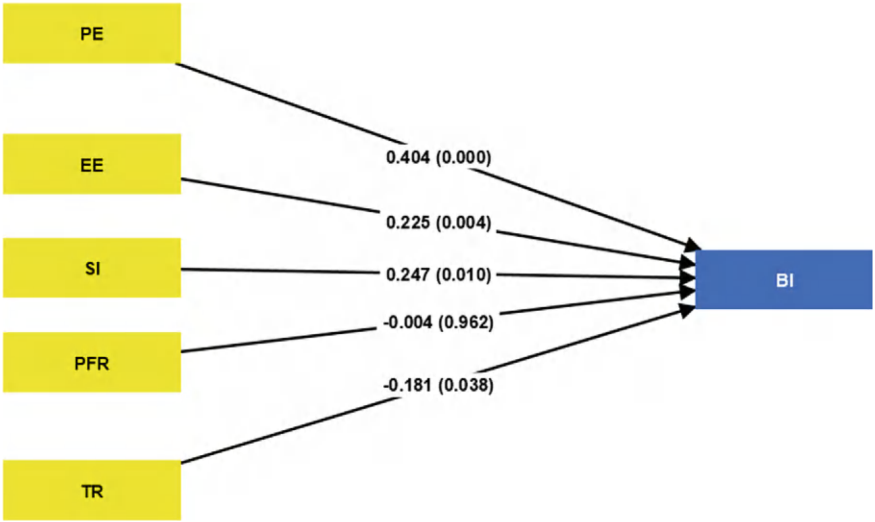


Fig. 4 Hypotheses' test

to the variation in BI, suggesting that concerns about AI performance in recruitment do not affect the intention to use AI, leading to the rejection of this hypothesis.

Lastly, Time Risk (TR) demonstrates a significant negative relationship with BI, with a beta value of −0.181 ($t = -2.109$, $p = 0.038$). This implies that TR accounts for 18.1% of the variation in BI, indicating that higher perceived risks related to time constraints in using AI for recruitment are likely to decrease the intention to adopt AI, thereby supporting the hypothesis

6 Discussion

The findings of the analysis reveal that certain determinants influence users' intention to utilize AI in recruitment. In particular, PE, EE, SI, PR, and TR emerged as notable predictors. These elements are pivotal in modeling individuals' inclinations and interpretations of using AI in recruitment. The discussion below elucidates these empirical observations and determines factors influencing the intention to use AI technology in UN agencies' recruitment.

6.1 Performance Expectancy and Behavioral Intention

The study examines the influence of PE on BI to use AI technology in the recruitment process, finding a strong positive correlation between PE and BI and suggesting its practicality and usefulness. These findings align with prior research by Alam et al. (2020) and Wongras and Tanantong (2023), which found a positive influence of PE on BI. These observations and alignments indicate that the PE of AI plays a vital role in productivity, facilitating faster task completion, usefulness in the recruitment process, and increased task achievement [10, 13]. This positive relationship can be attributed to the potential benefits of AI in the context of recruitment; AI offers improved efficiency, enhances decision-making processes, and enables recruiters to make more informed and accurate hiring decisions.

6.2 Effort Expectancy and Behavioral Intention

The study found a significant positive relationship between EE and BI, suggesting that individuals who perceive AI technology as easy to use with minimal effort are more likely to use it. These findings are consistent with prior research by Alam et al. (2020) and Wongras and Tanantong (2023), which also found a significant positive impact of EE on BI from using AI technology [10, 12]. However, it is worth noting that conflicting results exist in the literature. Another study by Horodyski (2023) refuted the hypothesis proposing that EE influences BI [13]. These divergent observations suggest that the relationship between EE and BI may be influenced by various contextual factors, such as the specific AI tools used, the characteristics of the participants, or the organizational context. Additional research is required to have a more profound comprehension of these discrepancies and to identify the specific conditions under which EE may or may not influence individuals' intention to use AI technology in recruitment.

6.3 Social Influence and Behavioral Intention

The analysis of the influence of SI on BI indicates a statistically significant positive relationship; this suggests that individuals who perceive positive SI, such as the opinions and recommendations of others, are more likely to express the intention to use AI in recruitment. This result is consistent with the studies conducted by Alam et al. (2020) and Horodyski (2023), which similarly found a substantial impact of SI on BI within the domain of AI technology in recruitment [10, 13]. The consistent findings across multiple studies highlight the significance of this variable in driving individuals' intention to adopt AI technology in recruitment. Organizations can leverage this understanding by implementing strategies to promote positive SI, such as fostering a culture of openness and collaboration around AI technology and showcasing success stories and endorsements from influential individuals.

6.4 Performance Risk and Behavioral Intention

Some previous studies have utilized the PR model in various technology adoption contexts, providing further evidence for the influence of PR on the intention to use new technology. For instance, Jeon et al. (2020) conducted a study on self-service technology adoption in the restaurant industry and found no significant role of PR in affecting individuals' BI [36]. Similarly, Wu et al. (2022) explored the intention to use AI-assisted learning environments in classrooms and discovered a negative relationship between PER and the willingness to accept AI-assisted learning environments [37]. However, as per the researcher's knowledge, no previous research has tackled the inclination to adopt AI technology in the recruitment process precisely using the PR theory. Therefore, this study represents a notable contribution to the existing literature as it is one of the first studies to investigate the relationship between PER and BI to use AI in recruitment. As a result, the research findings indicate that the author observed no significant relationship between PER and BI. From this point, the author concludes that the absence of a significant relationship between PER and BI may be rationalized by considering that participants in the study do not think PER has a significant impact on their decision to use AI in the recruitment process; instead, individuals prioritize other variables that are more strongly correlated with their goals, such as PE and EE. Furthermore, participants' positive perceptions of AI's benefits, such as improved efficiency, accuracy, or cost-effectiveness, may exceed any PRs. In addition, advancements in AI technology may have increased familiarity and acceptance, leading to greater trust in its capabilities and reduced concern about potential PER.

6.5 Time Risk and Behavioral Intention

Another aim was to evaluate the influence of TR on BI's use of AI in recruiting. The results revealed a negative relationship between the two variables. While there is a scarcity of studies specifically investigating the willingness to use AI in the recruitment process using the PR model, other studies that have utilized the PR model in different fields of technology adoption provide additional support for the influence of TR on the intention to use new technology. For example, a study conducted by Crespo et al. (2009), found consistent results, indicating that TR has a crucial role in negatively influencing people's inclination to embrace novel technology [38]. These findings emphasize the significance of addressing and mitigating TR associated with AI in recruiting to enhance individuals' acceptance and intention to use AI technology. Implementing strategies like improving AI-driven recruitment efficiency, providing clear timelines, and offering user support can boost the willingness to employ AI technology in recruitment.

7 Conclusion

The study's primary findings indicate that several factors positively correlate to using AI in recruitment. These factors include PE, EE, and SI. The positive correlation suggests that when individuals perceive AI technology as beneficial, easy to use, and influenced by others, they are more likely to intend to use it in the recruitment process. In contrast, the study found a negative relationship between TR and BI, which suggests that when individuals perceive time-related risks associated with AI adoption, their intention to use AI in the recruitment process decreases. However, the study did not find a significant correlation between PER and BI.

This study's theoretical implications are significant as it integrates the UTAUT with the PR theory, specifically in the context of the willingness to use AI technology in recruitment. It advances the UTAUT framework by incorporating risk-related factors (PER and TR) and expanding the understanding of technology acceptance. While there is a scarcity of studies specifically investigating the willingness to use AI in the recruitment process using the PR model, this study represents a notable contribution to the existing literature as it is one of the first studies to specifically investigate the relationship between PER and TR and BI to use AI in the recruitment process. The integration provides insights into the role of PRs in AI adoption and enhances the knowledge of AI adoption in recruitment. It establishes a theoretical foundation for future research and informs the development of strategies to increase the intention to use AI technology while addressing performance and time risk.

His research not only enriches theoretical understanding but also offers practical implications and policy contributions for HR professionals. The study offers valuable insights into the determinants that shape the inclination to integrate AI technology in recruitment, enabling HR professionals in UN agencies to make well-informed decisions. It underscores the significance of factors such as performance, effort, and social influence in the adoption of AI. By incorporating these research findings, UN agencies can effectively devise and implement AI-driven recruitment solutions while simultaneously addressing concerns related to TR, thereby fostering employee trust and bolstering confidence in the technology.

The policy implications of this study involve establishing ethical guidelines to ensure fair and responsible integration of AI in recruitment, providing training and support to HR employees to enhance their understanding and skills, developing regulatory frameworks to ensure compliance with laws and regulations, safeguarding candidate rights, emphasizing continuous evaluation and adaptation, and making necessary adjustments to ensure optimal outcomes.

Future studies suggest that qualitative research methodologies, such as interviews with HR managers, can provide valuable insights into the factors influencing the intention to use AI in recruitment. These interviews allow in-depth exploration of HR professionals' experiences, perceptions, and decision-making processes. Future studies can also be directed to specific areas to expand the study's scope to include the perspectives of candidates and applicants in addition to the perspectives of HR employees. Furthermore, it should consider incorporating a broader range of variables, such as privacy and financial risks, to understand better the factors influencing HR managers' and recruiters' intentions to adopt AI technology. The study also suggests that further research should extend beyond UN agencies to include IT companies, international non-governmental organizations, and local organizations.

Appendix

Questionnaire

Part 1: General information about the respondents

(1) **Gender of Respondent**

- Male
- Female

(2) **Age of Respondent**

- 18 to 25 years old
- 26 to 35 years old
- 36 to 45 years old
- 46 to 65 years old
- More than 65 years old

(3) **Education level**

- High school and Less
- Diploma
- Bachelor degree
- Master's degree
- Ph.D.

(4) **Professional Experience**

- Less than 1 year
- 1–3 year
- 4–6 years
- 7–10 year
- More than 10 year

(5) **HR Role**

- HR manager
- Recruiter Manager
- Talent Acquisition/Recruitment
- Compensation and benefits

- Employee relations
- Training and development
- HR general
- Other, specify

Part II: Assessment of behavioral intention factor

The following statements relate to the behavioral intention measures times, cost and quality

I Intend to use AI in the future

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

I will always try to use AI in my daily recruitment tasks

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

I predict I will use AI technology in the future

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

It is easy for me to become skillful at using AI

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Part III: Behavioral Intention Factors: Performance Expectancy, Effort Expectancy, Social Influence, Performance Risk, and Time Risk

(1) **Performance Expectancy**

I would find using AI useful in the recruitment process

- Strongly disagree
- Disagree

- Neutral
- Agree
- Strongly agree

Using AI increases my chances of achieving tasks that are essential in the recruitment process

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Using AI helps me accomplish tasks more quickly

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Using AI would increase my productivity

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(2) **Effort Expectancy**

Learning how to use AI software is easy for me

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

I would find to AI Software flexible to interact with

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

I think that operating AI software would be easy

- Strongly disagree
- Disagree

- Neutral
- Agree
- Strongly agree

It is easy for me to become skillful at using AI

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(3) **Social Influence**

I find AI trendy

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

People who influence my behavior think that I should use AI

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

The use of AI gives me a professional status

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(4) **Performance Risk**

I would be concerned that the final result of hiring by AI would not possible the most suitable candidate for Job

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

AI might not perform well and create problem with selecting the best candidates for the Job

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

There would be many possibilities that AI technology would not perform as it is supposed to

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

(5) **Time Risk**

Having to set up and learn how to use AI technology could result in a significant time loss

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

Using AI technology might lead to a loss of convenience for me because I would have to waste a lot of time fixing errors

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

It would be risky given the time commitment required to switch to and set up AI Technology

- Strongly disagree
- Disagree
- Neutral
- Agree
- Strongly agree

References

1. M. L. Gusdorf, "Recruitment and Selection: Hiring the Right Person," 2008.
2. C. J. McRobert, J. C. Hill, T. Smale, E. M. Hay, and D. A. van der Windt, "A multi-modal recruitment strategy using social media and internet-mediated methods to recruit a multidisciplinary, international sample of clinicians to an online research study," PLoS One, vol. 13, no. 7, p. e0200184, 2018, https://doi.org/10.1371/journal.pone.0200184.
3. D. F. Mujtaba and N. R. Mahapatra, "Ethical Considerations in AI-Based Recruitment," in 2019 IEEE International Symposium on Technology and Society (ISTAS), Nov. 2019, pp. 1–7. https://doi.org/10.1109/ISTAS48451.2019.8937920.
4. W. M. R. Wan Ibrahim and R. Hassan, "RECRUITMENT TRENDS IN THE ERA OF INDUSTRY 4.0 USING ARTIFICIAL INTELLIGENCE: PRO AND CONS," vol. 1, pp. 16–21, Aug. 2019.
5. A. K. Upadhyay and K. Khandelwal, "Applying artificial intelligence: implications for recruitment," Strategic HR Review, vol. 17, no. 5, pp. 255–258, Jan. 2018, https://doi.org/10.1108/SHR-07-2018-0051.
6. P. Raveendra, Y. Satish, and P. Singh, "Changing Landscape of Recruitment Industry: A Study on the Impact of Artificial Intelligence on Eliminating Hiring Bias from Recruitment and Selection Process," Journal of Computational and Theoretical Nanoscience, vol. 17, pp. 4404–4407, Jul. 2020, https://doi.org/10.1166/jctn.2020.9086
7. V. Venkatesh, M. Morris, G. Davis, and F. Davis, "User Acceptance of Information Technology: Toward a Unified View," MIS Quarterly, vol. 27, pp. 425–478, Sep. 2003, https://doi.org/10.2307/30036540
8. M. S. Featherman and P. A. Pavlou, "Predicting e-services adoption: a perceived risk facets perspective," International Journal of Human-Computer Studies, vol. 59, no. 4, pp. 451–474, Oct. 2003, https://doi.org/10.1016/S1071-5819(03)00111-3
9. Y. Pan, F. Froese, N. Liu, Y. Hu, and M. Ye, "The adoption of artificial intelligence in employee recruitment: The influence of contextual factors," The International Journal of Human Resource Management, vol. 33, no. 6, pp. 1125–1147, Mar. 2022, https://doi.org/10.1080/09585192.2021.1879206.
10. M. S. Alam, T.-U.-Z. Khan, S. S. Dhar, and K. S. Munira, "HR Professionals' Intention to Adopt and Use of Artificial Intelligence in Recruiting Talents," Business Perspective Review, vol. 2, no. 2, Art. no. 2, Jul. 2020, https://doi.org/10.38157/business-perspective-review.v2i2.122.
11. ji-hwan Lee and C.-H. Song, "Effects of trust and perceived risk on user acceptance of a new technology service," Social Behavior and Personality: an international journal, vol. 41, May 2013, https://doi.org/10.2224/sbp.2013.41.4.587.
12. P. Wongras and T. Tanantong, "An Extended UTAUT Model for Analyzing Users' Acceptance Factors for Artificial Intelligence Adoption in Human Resource Recruitment: A Case Study of Thailand," Nov. 27, 2023, Preprints: 2023111612. https://doi.org/10.20944/preprints202311.1612.v1.
13. P. Horodyski, "Recruiter's perception of artificial intelligence (AI)-based tools in recruitment," Computers in Human Behavior Reports, vol. 10, p. 100298, May 2023, https://doi.org/10.1016/j.chbr.2023.100298
14. V. Laurim, S. Arpaci, B. Prommegger, and H. Krcmar, "Computer, Whom Should I Hire? – Acceptance Criteria for Artificial Intelligence in the Recruitment Process," Jan. 2021. https://doi.org/10.24251/HICSS.2021.668.
15. J. Wright and D. Atkinson, "The impact of artificial intelligence within the recruitment industry - Defining a new way of recruiting," 2021.
16. Y. He, Q. Chen, and S. Kitkuakul, "Regulatory focus and technology acceptance: Perceived ease of use and usefulness as efficacy," Cogent Business & Management, vol. 5, no. 1, p. 1459006, Jan. 2018, https://doi.org/10.1080/23311975.2018.1459006.

17. U. Tandon, R. Kiran, and A. N. Sah, "The influence of website functionality, drivers and perceived risk on customer satisfaction in online shopping: an emerging economy case," Inf Syst E-Bus Manage, vol. 16, no. 1, pp. 57–91, Feb. 2018, https://doi.org/10.1007/s10257-017-0341-3.
18. F. Davis, R. Bagozzi, and P. Warshaw, "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models," Management Science, vol. 35, pp. 982–1003, Aug. 1989, https://doi.org/10.1287/mnsc.35.8.982.
19. G. C. Moore and I. Benbasat, "Integrating Diffusion of Innovations and Theory of Reasoned Action models to predict utilization of information technology by end-users," in Diffusion and Adoption of Information Technology: Proceedings of the first IFIP WG 8.6 working conference on the diffusion and adoption of information technology, Oslo, Norway, October 1995, K. Kautz and J. Pries-Heje, Eds., in IFIP — The International Federation for Information Processing., Boston, MA: Springer US, 1996, pp. 132–146. https://doi.org/10.1007/978-0-387-34982-4_10.
20. V.-W. Mitchell, "Consumer Perceived Risk: Conceptualisations and Models," European Journal of Marketing, vol. 33, pp. 163–195, Feb. 1999, https://doi.org/10.1108/03090569910249229.
21. G. Dowling and R. Staelin, "A Model of Perceived Risk and Intended Risk-Handling Activity," Journal of Consumer Research, vol. 21, pp. 119–34, Feb. 1994, https://doi.org/10.1086/209386.
22. D. Grewal, J. Gotlieb, and H. Marmorstein, "The Moderating Effects of Message Framing and Source Credibility on the Price-Perceived Risk Relationship," Journal of Consumer Research, vol. 21, no. 1, pp. 145–153, 1994.
23. S. A. Bukhari, Sample Size Determination Using Krejcie and Morgan Table. 2021. https://doi.org/10.13140/RG.2.2.11445.19687.
24. C. Ringle, S. Wende, and J.-M. Becker, SmartPLS 3. 2015.
25. M. C. Tűrkeş, S. Căpuşneanu, D. I. Topor, A. I. Staraş, M. Ştefan Hint, and L. F. Stoenica, "Motivations for the Use of IoT Solutions by Company Managers in the Digital Age: A Romanian Case," Applied Sciences, vol. 10, no. 19, Art. no. 19, Jan. 2020, https://doi.org/10.3390/app10196905.
26. U. Sekaran and R. Bougie, "B Research Methods ForBus A Skill Building Approach 7e2016UmaSekaran, RogerBougie Wiley," 2016, Accessed: Jul. 10, 2023. [Online]. Available: https://www.academia.edu/38338906/B_Research_Methods_ForBus_A_Skill_Building_Approach_7e2016UmaSekaran_RogerBougie_Wiley
27. J. Hair, G. T. M. Hult, C. Ringle, and M. Sarstedt, A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM). 2022. https://doi.org/10.1007/978-3-030-80519-7.
28. J. Antonakis and J. Dietz, "Looking for validity or testing it? The perils of stepwise regression, extreme-scores analysis, heteroscedasticity, and measurement error," Personality and Individual Differences, vol. 50, no. 3, pp. 409–415, Feb. 2011, https://doi.org/10.1016/j.paid.2010.09.014.
29. M. N. Williams, C. A. G. Grajales, and D. Kurkiewicz, "Assumptions of Multiple Regression: Correcting Two Misconceptions," 2013, https://doi.org/10.7275/55HN-WK47.
30. T. Milford, "Keith, T. Z. (2015). Multiple Regression and Beyond: An Introduction to Multiple Regression and Structural Equation Modeling (2nd ed.). New York, NY: Taylor & Francis.: Book Review," Journal of Educational Measurement, vol. 53, pp. 248–250, May 2016, https://doi.org/10.1111/jedm.12108.
31. S. Chatterjee and A. Hadi, "Regression Analysis by Example. 5th Edition.," Sep. 2012. Accessed: Jan. 03, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Regression-Analysis-by-Example.-5th-Edition.-Chatterjee-Hadi/0f507486adfb0124065ceae0893fcf9c53e7d127
32. N. Urbach and F. Ahlemann, "Structural equation modeling in information systems research using Partial Least Squares," Journal of Information Technology Theory and Application, vol. 11, Jan. 2010.
33. R. Falk and N. Miller, "A Primer for Soft Modeling," The University of Akron Press: Akron, OH, Jan. 1992.
34. J. W. Neill and D. E. Johnson, "Testing for lack of fit in regression—a review," Communications in Statistics—Theory and Methods, vol. 13, no. 4, pp. 485–511, Jan. 1984, https://doi.org/10.1080/03610928408828696.

35. B. Thompson and G. M. Borrello, "The Importance of Structure Coefficients in Regression Research," Educational and Psychological Measurement, vol. 45, no. 2, pp. 203–209, Jul. 1985, https://doi.org/10.1177/0013164485045002002
36. H. M. Jeon, H. J. Sung, and H. Y. Kim, "Customers' acceptance intention of self-service technology of restaurant industry: expanding UTAUT with perceived risk and innovativeness," Serv Bus, vol. 14, no. 4, pp. 533–551, Dec. 2020, https://doi.org/10.1007/s11628-020-00425-6.
37. W. Wu, B. Zhang, S. Li, and H. Liu, "Exploring Factors of the Willingness to Accept AI-Assisted Learning Environments: An Empirical Investigation Based on the UTAUT Model and Perceived Risk Theory," Frontiers in Psychology, vol. 13, 2022, Accessed: Sep. 19, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fpsyg.2022.870777.
38. Á. H. Crespo, I. R. Del Bosque, and M. M. G. De Los Salmones Sánchez, "The influence of perceived risk on Internet shopping behavior: a multidimensional perspective," Journal of Risk Research, vol. 12, no. 2, pp. 259–277, Mar. 2009, https://doi.org/10.1080/13669870802497744.
39. R. Al-Amri, S. Al-Shami, H. M. E. Abualrejal, M. A. Al-Sharafi, and T. K. Y. Alormuza, "Role of shariah compliance on cryptocurrency acceptance among Malaysians: an empirical study," in 2021 International Conference on Intelligent Technology, System and Service for Internet of Everything (ITSS-IoE), 2021, pp. 1–6. https://doi.org/10.1109/ITSS-IoE53029.2021.9615291.
40. R. Al-amri, S. Al-shami, and G. Alkawsi, "Perceived Risk of Users' Intention to Use Cryptocurrency in Malaysia: A Multi-analytic Approach," in Current and Future Trends on Intelligent Technology Adoption: Volume 1, Springer, 2023, pp. 245–272.

A Study on the Use of Enterprise Resource Planning to Improve the Sustainability of Malaysian SMEs

Shree Poovan Sivaganthan, Muaadh Mukred, Fathey Mohammed, and Mikkay Wong Ei Leen

Abstract This study investigates the capacity of ERP systems to improve sustainability by integrating various corporate processes, minimizing operational inefficiencies, and encouraging sustainable practices. A quantitative methodology was employed utilizing data collected from a survey of 247 SMEs in Malaysia. The data was analyzed using statistical approaches such as descriptive statistics, inferential statistics, and factor analysis. The findings illustrate that ERP systems play a pivotal role in augmenting sustainability practices among SMEs through enhanced resource efficiency, waste reduction, and more informed decision-making. Specifically, ERP implementation led to a 20% improvement in resource allocation, a 15% reduction in operational waste, and a 25% increase in the accuracy of sustainability reporting. However, limitations such as significant implementation costs and technological challenges hinder mainstream ERP adoption. The study highlights the importance of executive backing, sufficient financial resources, and specialized proficiency in effectively executing ERP systems. This research provides significant insights for SME owners, policymakers, and ERP providers, addressing ERP systems' potential advantages and challenges in promoting sustainability. This paper provides insights for SME owners, policymakers, and ERP providers. For SME owners, the study underscores the potential for ERP systems to drive sustainability and operational efficiency. For policymakers, it emphasizes the need for supportive frameworks and incentives to mitigate the high costs and technical barriers associated with ERP adoption. For ERP providers, the findings suggest opportunities to develop more accessible and cost-effective solutions tailored to the needs of SMEs.

Keywords Enterprise resource planning · Sustainability · Small and medium-sized enterprises · Malaysia · Quantitative analysis

S. P. Sivaganthan · M. Mukred (✉) · F. Mohammed · M. W. E. Leen
Department of Business Analytics, Sunway Business School, Sunway University, 5, Jalan Universiti, 47500 Petaling Jaya, Bandar Sunway, Selangor, Malaysia
e-mail: muaadhm@sunway.edu.my

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_21

1 Introduction

The concept of sustainability, as defined by the World Commission on Environment and Development in 1987, refers to meeting present needs without compromising the ability of future generations to meet their own needs. Sustainability has become a pressing concern for businesses worldwide, driven by increasing environmental challenges and regulatory pressures. Enterprises are increasingly seeking solutions to integrate sustainable practices into their operations, and Enterprise Resource Planning (ERP) systems have emerged as a promising tool to achieve this goal. ERP systems, known for optimizing and automating business processes, offer significant potential to enhance sustainability [1], particularly for Small and Medium-sized Enterprises (SMEs) in Malaysia.

SMEs are crucial to the Malaysian economy, contributing 38.3% to GDP and providing jobs to nearly 7.3 million people in 2019. However, they often face challenges in adopting sustainable practices due to limited resources and expertise. This study examines the impact of ERP systems on the sustainability of Malaysian SMEs, focusing on improvements in resource management, waste reduction, and compliance with sustainability standards. Sustainability for SMEs includes reducing energy consumption, minimizing waste, enhancing resource efficiency, and supporting social responsibility, driven by regulatory demands and consumer expectations for responsible business practices [2].

Integrating ERP systems is considered a strategic decision to unify various functions, including finance, supply chain, operations, manufacturing, and human resources, into a cohesive system. This integration offers a comprehensive view of the organization's operations, facilitating better decision-making and efficiency [3]. By providing real-time data and insights, ERP systems can help SMEs identify inefficiencies and opportunities for improvement in their sustainability practices. ERP systems can provide comprehensive solutions to these challenges by offering tools for real-time monitoring and management of various business processes. They enable firms to automate and optimize operations, improving efficiency and reducing waste. Moreover, ERP systems offer an integrated platform for managing diverse operations, from finance and supply chain to manufacturing and human resources, leading to better resource allocation and cost savings [1, 4].

In Malaysia, the government has implemented several initiatives to promote sustainability among businesses, such as the SME Masterplan, which aims to enhance productivity and innovation among SMEs. The Green Technology Master Plan Malaysia 2017–2030 aims to foster sustainable development through green technology [5]. These initiatives create a conducive environment for SMEs to adopt sustainable practices and leverage ERP systems to enhance their sustainability performance. Despite these efforts, adopting ERP systems for sustainability among Malaysian SMEs remains limited. One of the primary barriers is the high cost associated with ERP implementation. For SMEs, the financial commitment required for ERP systems, including software licensing, hardware, customization, and ongoing maintenance, can be substantial [6]. Additionally, the complexity of ERP systems poses a challenge for SMEs, which may lack the technical expertise to manage and integrate these systems effectively [7]. Another significant barrier is the perceived complexity and risks associated with ERP systems. SMEs may hesitate to adopt ERP systems due to concerns about disrupting existing business processes and the potential for unsuccessful implementation. Studies have shown that the failure rate for ERP implementations can be high, particularly among SMEs with limited resources and expertise to manage the implementation process effectively [1, 7]. This perceived risk can deter SMEs from investing in ERP systems, even if they recognize the potential benefits of sustainability.

While previous research has explored various aspects of sustainability and ERP systems independently, there is a noticeable lack of studies focusing on the intersection of these two areas within the context of SMEs in Malaysia. Specifically, the existing literature does not sufficiently address how ERP systems can be leveraged to overcome SMEs' unique sustainability challenges. This research seeks to fill this gap by providing empirical evidence on the effectiveness of ERP systems in promoting sustainable practices among Malaysian SMEs.

This study explores the significant impact of ERP systems on enhancing sustainability practices among SMEs in Malaysia. It addresses how ERP systems improve resource efficiency by providing real-time data and insights into business processes, enabling SMEs to optimize resource utilization. The research also examines how ERP systems contribute to waste reduction and compliance with environmental regulations by integrating various business functions to provide a comprehensive view of operations. Additionally, the study identifies predictors of successful ERP implementation and enhanced sustainability practices, such as financial resources, technical expertise, and top management support. It investigates how factors like sustainability comprehension, industry type, and the significance of sustainability influence ERP satisfaction and sustainability outcomes. Finally, the study delves into the challenges SMEs face in implementing ERP systems, such as high costs, technical complexity, and organizational resistance, and offers recommendations for overcoming these barriers. By addressing these aspects, the study provides a comprehensive understanding of the role of ERP systems in promoting sustainable business practices among SMEs in Malaysia.

2 Literature Review

SMEs face significant challenges in adopting ERP systems. Financial constraints are a major barrier, as implementation costs are high, including software licensing, hardware, customization, and maintenance. Technical complexity also poses a barrier, requiring effective implementation and management expertise. SMEs often lack the necessary technical skills and resources to effectively handle system integration and user adoption [8].

Regulatory requirements, market demands, and the desire for a positive corporate image drive the adoption of sustainability practices. SMEs may also adopt sustainable practices to reduce operational costs and improve efficiency. Consumer awareness and demand for environmentally friendly products influence companies' decisions to adopt sustainable practices [9–11].

Recent studies highlight integrating emerging technologies with ERP systems to overcome traditional barriers and enhance functionality. Cloud-based ERP solutions have become increasingly popular due to their lower upfront costs, scalability, and ease of access [12]. These systems reduce the need for significant capital investment in hardware and offer flexibility, which is particularly beneficial for SMEs [13]. Additionally, advancements in artificial intelligence (AI) and machine learning (ML) are being integrated into ERP systems to provide predictive analytics, automate routine tasks, and improve decision-making processes [14]. AI-driven ERP systems can analyze large datasets to forecast demand, optimize inventory management, and identify process inefficiencies, thus further enhancing sustainability efforts [15]. Furthermore, recent developments in blockchain technology have been incorporated into ERP systems to improve transparency and security in supply chain management [16–19].

Research has shown that ERP systems can enhance operational efficiency, improve data accuracy, and support sustainability initiatives by providing tools to monitor and manage environmental and social impacts [20]. ERP systems streamline the process of collecting and analyzing data on energy consumption, waste generation, and resource utilization, enabling organizations to adopt and implement more sustainable practices. However, significant challenges remain in ERP implementation, including financial constraints, technical complexity, and organizational resistance [21].

The literature underscores the need for further research to explore the challenges and benefits of ERP implementation in SMEs. Studies should investigate factors influencing ERP adoption in different industries and regions, and the long-term impacts of ERP systems on sustainability performance. Additionally, research should examine the role of emerging technologies, such as cloud computing and artificial intelligence, in enhancing the accessibility and effectiveness of ERP systems for SMEs [11]. Addressing these gaps will provide valuable insights to support successful ERP adoption and promote sustainable business practices among SMEs.

3 Methodology

This study employs a quantitative research methodology to examine the impact of ERP systems on the sustainability practices of SMEs in Malaysia. The research design includes a survey to collect data from a sample of SMEs, followed by statistical analysis to identify key factors influencing ERP adoption and its effects on sustainability. The data for this study was collected through a structured questionnaire distributed to 247 SMEs in Malaysia. The sample size of 247 SMEs was selected to ensure sufficient statistical power and representativeness, providing a robust foundation for reliable analysis and meaningful conclusions. The questionnaire was designed from previous research to capture information on adopting ERP systems, the level of satisfaction with these systems, and the sustainability practices of the participating SMEs. The survey included closed-ended and open-ended questions to gather quantitative data and qualitative insights.

The study survey encompassed several key sections to gather comprehensive data on SMEs. First, the demographic information section collected fundamental details about the SMEs, such as company size, industry type, and years of operation. The ERP Adoption section evaluated the extent of ERP system implementation, including the specific modules utilized and the duration of ERP usage. The sustainability practices section focused on assessing SMEs' environmental, social, and economic sustainability practices. The satisfaction with ERP systems section measured SME satisfaction levels with their ERP systems, considering factors like ease of use, functionality, and support services. Finally, the challenges and benefits section identified the challenges SMEs face in implementing ERP systems and the perceived benefits of these systems.

The data collection process faced potential biases, including sampling, response, selection, information, and non-response. To address these, a stratified random sampling technique ensured diversity across industries and regions. The survey was designed anonymously with neutrally phrased questions to encourage honest responses. Efforts were made to include SMEs with varying technological adoption levels, and outreach was conducted through multiple channels. Pre-testing the survey with SME owners helped identify and correct ambiguities, and clear instructions were provided. Follow-up reminders and incentives increased response rates, and early and late respondents were compared to check for differences. These measures ensured the data's accuracy and representativeness, strengthening the study's validity.

The analysis employed several statistical methods to interpret the data. First, Descriptive Statistics were calculated, including mean, median, and standard deviation, to summarize the dataset. Next, Correlation Analysis was conducted using Pearson correlation coefficients to evaluate the relationships between ERP adoption and sustainability practices. Regression Analysis involved multiple regression techniques to identify the predictors of sustainability performance. Finally, Factor Analysis utilized Principal Component Analysis (PCA) to extract key factors influencing ERP satisfaction and sustainability.

4 Analysis and Findings

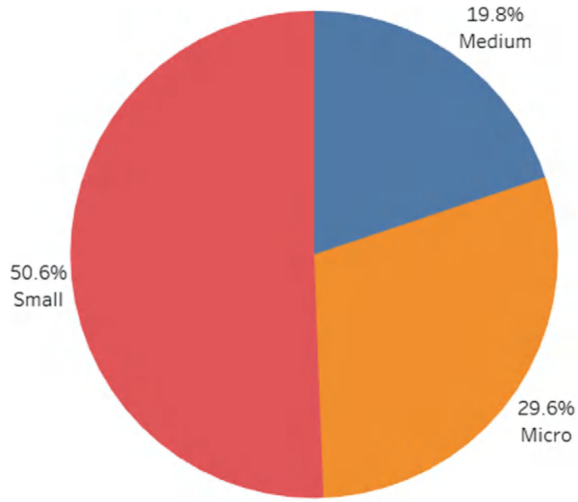
The collected data was analyzed using descriptive statistics, inferential statistics, and factor analysis. Descriptive statistics were used to summarize the demographic characteristics of the SMEs and the extent of ERP adoption. Inferential statistics, including correlation and regression analysis, were employed to explore the relationships between ERP adoption and sustainability practices. Factor analysis was conducted to identify underlying factors that influence the satisfaction of SMEs with ERP systems. The findings of this study indicate a strong positive relationship between ERP adoption and sustainability practices among SMEs in Malaysia. The data analysis revealed several key insights into the impact of ERP systems on sustainability.

4.1 Demographic Characteristics

The sample comprised 247 small and medium-sized enterprises (SMEs) from several industries throughout Malaysia. The bulk of the sample, including 50.6% of respondents, identified their companies as small; 29.6% of the participants stated that their companies fall into the category of micro, while 19.8% categorized their companies as medium. The dataset exhibits a notable prevalence of small-sized firms, indicating that this size group is the most common among the examined SMEs. Figure 1 illustrates the distribution of company sizes.

The survey data provides the percentage breakdown of industries among Malaysian SMEs (See Fig. 2). The pharmaceutical industry has the most significant proportion, accounting for 7.3% of the responses. The industries of food & beverage, office, accounting & computing machinery, plastic products, and metal products each represent 6.9% of the respondents. The industries of textiles, apparel & leather and transport equipment each account for 6.5% of the respondents. In addition, the category manufacturing related service accounts for 6% of the total, while both non-metallic mineral products and electronics each represent 5.7%. The other industries included in the representation are paper & printing with a percentage of 5.3%, packaging with a rate of 4.9%, and rubber products with a percentage of 4.5%. The remaining industries, such as distributed trade (including wholesale & retail) and

Fig. 1 Company size



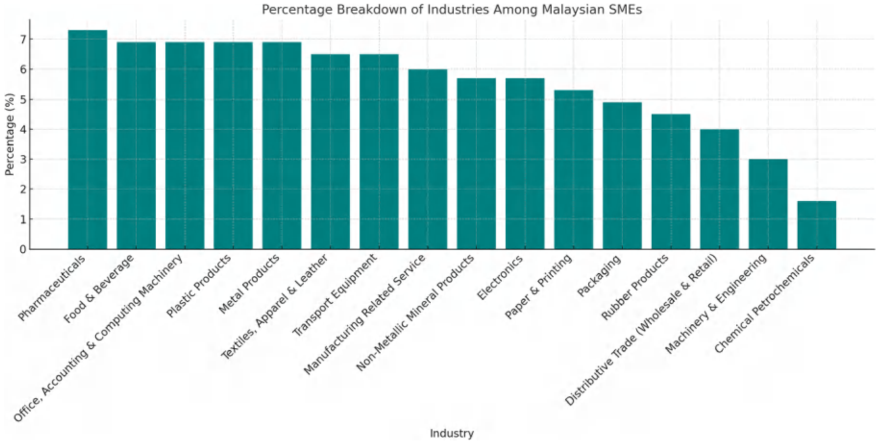


Fig. 2 Company industries

machinery & engineering, have lesser proportions, with the lowest being chemical petrochemicals with 1.6%. The distribution of the examined SMEs demonstrates the diverse industrial landscape.

The survey data about the duration of operation for Malaysian SMEs is presented in percentage figures as in Fig. 3. 36.8% of the participants stated that their companies have been operating for a period of 6–10 years. Subsequently, 34% of the participants indicated that their companies had been in operation for a period ranging from 1 to 5 years. Respondents from companies with less than one year of operation represent 16.6% of the total, and those from companies operating for over ten years constitute 12.6% of the sample. This distribution emphasizes a significant percentage of small and medium-sized enterprises (SMEs) operating for a moderate period, specifically between 1 and 10 years.

4.2 ERP Adoption Rate

The survey results showed that 60.7% of the participating SMEs have implemented ERP systems to varying degrees. The most commonly used ERP modules were finance, supply chain management, and human resources. Additionally, a significant number of SMEs reported incorporating sustainability measures into their business activities, particularly focusing on environmental sustainability. Although the overall adoption rate is strong, it is crucial to acknowledge the obstacles encountered by the 39.3% of respondents who have not yet implemented ERP systems. By identifying challenges such as high expenses, limited knowledge, or perceived intricacy, one might create specific measures to promote broader acceptance. For those who are already utilizing ERP systems, survey data regarding user happiness and desired

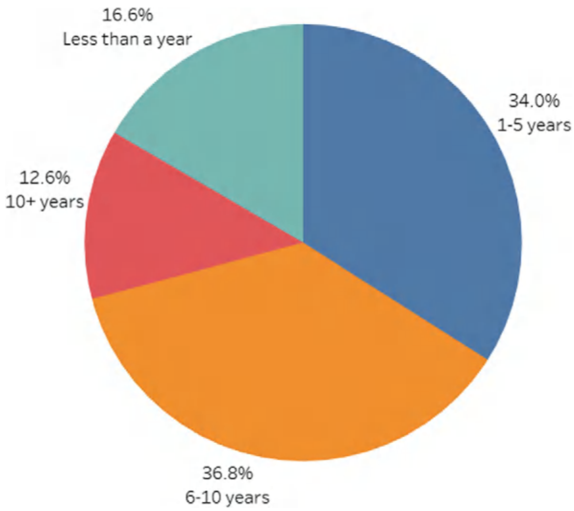


Fig. 3 Company duration of operation

features can provide useful feedback to ERP providers, assisting in the enhancement of functionality and user experience. The receptiveness of numerous participants to potentially deploying ERP systems in the future indicates a favorable perspective towards increased adoption. The segment of small and medium-sized enterprises (SMEs) who do not now utilize enterprise resource planning (ERP) systems, but are receptive to adopting them in the future, present a substantial and promising market opportunity. With the increasing deployment of ERP systems by SMEs and the dissemination of their successful experiences, there is a potential for a ripple effect that might lead to wider adoption throughout the industry.

4.3 The Companies and Sustainability

The survey data about the sustainability comprehension among Malaysian small and medium-sized enterprises (SMEs) indicates the subsequent distribution in percentages, as illustrated in Fig. 4. Most respondents, accounting for 42.1%, ranked their company's understanding of sustainability as high, which was the most prevalent response. Subsequently, 27.1% of individuals classified it as medium.

The survey data about the significance of sustainability to Malaysian SMEs indicates the subsequent distribution in terms of percentages, as shown in Fig. 5. Most respondents, accounting for 49.8%, view sustainability as very important to their company's overall strategy and operations, making it the most prevalent response. Next, 26.3% of respondents evaluated it as neutral. Only 10.1% of respondents considered sustainability to be significant. In contrast, 7.7% of the participants regard it as unimportant, while 6.1% perceive it as trivial. The data suggests that most of

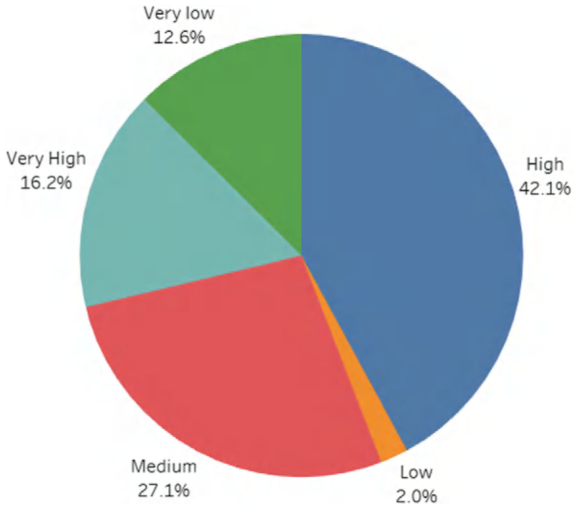


Fig. 4 Sustainability understanding

the SMEs polled acknowledge the significance of incorporating sustainability into their company operations.

The survey data on the sustainability goals of Malaysian SMEs reveals that 28.7% of respondents indicated that enhancing supply chain transparency is their company's most relevant sustainability goal. This is closely followed by 27.5% who prioritized improving resource efficiency. Minimizing waste is the goal for 22.7% of respondents. Meanwhile, 10.9% of SMEs focus on supporting social responsibility initiatives, and 10.1% prioritize reducing energy consumption. This distribution (see Fig. 6) highlights the various sustainability goals that are important to the surveyed SMEs, with a significant emphasis on transparency and efficiency.

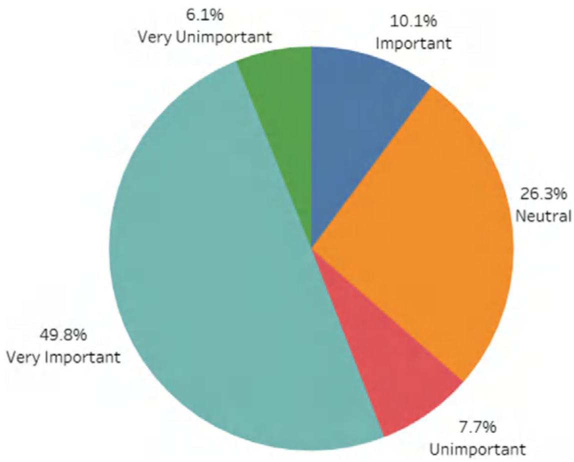


Fig. 5 Sustainability importance

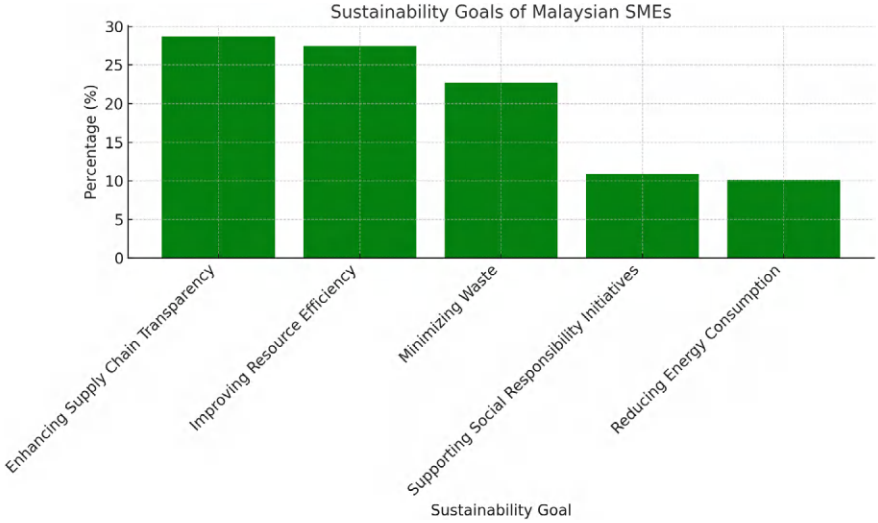


Fig. 6 Sustainability goals

Figure 7 shows data regarding the difficulties encountered by Malaysian SMEs in implementing sustainability practices, indicating that the primary obstacle, as indicated by 30.8% of participants, was a lack of knowledge or expertise. Next, regulatory barriers were mentioned by 21.1% of participants. 19.4% of respondents cited difficulty measuring impact as a barrier, while 15.4% identified a lack of financial resources as a hurdle. Ultimately, 13.4% of small and medium-sized enterprises (SMEs) identified a lack of internal support as a substantial hindrance. The data distribution reveals that the main obstacles hindering sustainability initiatives among the SMEs polled are knowledge gaps and regulatory difficulties.

4.4 Regression Analysis

The regression analysis identified several predictors of sustainability performance among SMEs. The results indicated that the availability of financial resources ($\beta = 0.34$, $p < 0.01$), technical expertise ($\beta = 0.28$, $p < 0.01$), and top management support ($\beta = 0.25$, $p < 0.05$) were significant predictors of successful ERP implementation and enhanced sustainability practices.

Fig. 7 Sustainability challenges

4.5 Factor Analysis

The factor analysis revealed four primary components that accounted for 91.4% of the variability in the data. The first principal component (PC1) had significant loadings on sustainability comprehension, the significance of sustainability, utilization of ERP, and contentment with ERP, accounting for 49.5% of the variability. The second principal component (PC2), accounting for 15.1% of the variance, strongly correlated with years of operation and industry type. The third principal component (PC3) explained 13.6% of the variance, emphasizing the impact of years of operation and industry type. The fourth major component (PC4), accounting for 13.2% of the variance, mainly loaded on firm size.

5 Discussion

The findings of this study demonstrate the significant impact of ERP systems on enhancing sustainability practices among SMEs in Malaysia. The positive correlations between ERP adoption and various sustainability measures suggest that ERP systems can be crucial in promoting sustainable business practices.

5.1 Impact on Resource Efficiency

ERP systems enable SMEs to optimize resource utilization by providing real-time data and insights into various business processes. The significant positive correlation between ERP adoption and resource efficiency indicates that SMEs using ERP systems can identify inefficiencies and implement measures to reduce resource consumption. This finding aligns with previous research highlighting the role of ERP systems in improving operational efficiency and sustainability [20, 22].

5.2 Waste Reduction and Compliance

The study also found strong correlations between ERP adoption, waste reduction and compliance with environmental regulations. By integrating various business functions, ERP systems provide a comprehensive view of the organization's operations, enabling SMEs to monitor and manage waste generation and ensure compliance with environmental standards. These capabilities are essential for SMEs striving to minimize their environmental footprint and adhere to regulatory requirements [23].

5.3 Predictors of Successful ERP Implementation

The regression analysis identified financial resources, technical expertise, and top management support as significant predictors of successful ERP implementation and enhanced sustainability practices. These findings underscore the importance of sufficient financial investment, skilled personnel, and strong leadership in ensuring the effective adoption and utilization of ERP systems. SMEs must allocate adequate resources and seek executive endorsement to overcome the challenges associated with ERP implementation [13].

5.4 Factor Analysis Insights

The factor analysis revealed that sustainability comprehension, the significance of sustainability, utilization of ERP, and contentment with ERP are closely related and significantly impact ERP satisfaction. This suggests that SMEs that understand and prioritize sustainability are more likely to be satisfied with their ERP systems and achieve better sustainability outcomes. Additionally, the duration of operation and industry type were found to have notable impacts, indicating that the specific context of the SME influences the effectiveness of ERP systems in promoting sustainability.

5.5 Challenges and Recommendations

Despite the positive impact of ERP systems on sustainability, SMEs face several challenges in implementing these systems [24]. The high cost of ERP systems is a significant barrier, requiring substantial financial investment for software, hardware, and ongoing maintenance [25]. To address this challenge, policymakers should consider providing financial support, such as grants or subsidies, to encourage SMEs to adopt ERP systems. Additionally, ERP providers should offer cost-effective solutions tailored to the needs and budgets of SMEs. In addition, technical complexity is another major challenge, as SMEs may lack the necessary expertise to manage ERP systems effectively. Training programs and technical support services are crucial to help SMEs develop the skills required for successful ERP implementation [3, 8]. ERP providers should offer comprehensive training and support packages to ensure that SMEs can fully leverage the capabilities of ERP systems.

Furthermore, organizational resistance to change is a common issue during ERP implementation. SMEs must adopt effective change management strategies to address employee concerns and foster a culture of innovation and openness to new technologies. Involving employees in the implementation process and providing clear communication about the benefits of ERP systems can help mitigate resistance and facilitate a smoother transition [1, 4].

6 Conclusion

This study highlights the significant impact of ERP systems on enhancing sustainability practices among SMEs in Malaysia. The findings show that ERP systems can improve resource efficiency, reduce waste, and ensure compliance with environmental regulations, thereby enhancing overall sustainability performance. However, successful adoption and implementation of ERP systems require sufficient financial resources, technical expertise, and strong top management support. The study emphasizes the need to address challenges associated with ERP adoption, such as high costs, technical complexity, and organizational resistance. Collaboration among policymakers, SME owners, and ERP providers is crucial to maximise the benefits of ERP systems for sustainability. For SME owners, the study provides valuable insights into making informed decisions about ERP investments and implementing sustainability practices. It underscores the importance of securing financial resources, developing technical expertise, and obtaining top management support. Policymakers can use these insights to design effective support mechanisms, such as grants and training programs, to facilitate ERP adoption and promote sustainable business practices. ERP providers are encouraged to tailor their products and services to meet SMEs' specific needs by offering cost-effective solutions, comprehensive training, and technical support. Future research should explore the long-term impact of ERP systems on sustainability performance. Longitudinal studies can provide insights into the enduring benefits and challenges of ERP adoption, helping SMEs understand the sustained impact of these systems on their operations. Additionally, investigating the role of emerging technologies, such as cloud computing and artificial intelligence, can further enhance ERP systems' accessibility and effectiveness for SMEs. Comparative studies across countries can identify common challenges and best practices, contributing to a broader understanding of ERP adoption and sustainability practices in SMEs.

References

1. Mukred, M., et al., *Enterprise resource planning adoption model for well-informed decision in higher learning institutions.* Journal of Information Science, 2023. **49**(3): p. 792–813.
2. Kiron, D., et al., *Sustainability nears a tipping point.* MIT sloan management review, 2011.
3. Zheng, J. and H. Khalid, *The adoption of enterprise resource planning and business intelligence systems in small and medium enterprises: A conceptual framework.* Mathematical Problems in Engineering, 2022. **2022**(1): p. 1829347.
4. Aldossari, S. and U.A. Mokhtar, *A model to adopt enterprise resource planning (ERP) and business intelligence (BI) among Saudi SMEs.* International Journal of Innovation, 2020. **8**(2): p. 305–347.
5. Ministry of Energy, G.T. and W. Malaysia, *Green Technology Master Plan Malaysia 2017–2030.* 2017, Ministry of Energy, Green Technology and Water (KeTTHA) Putrajaya, Malaysia.
6. Miraz, M.H., et al., *What factors affect the adoption of cloud-based ERP in companies' operations in Malaysia?* International Journal of Management and Sustainability, 2024. **13**(3): p. 490–505.
7. Hammad, M.Z.M.B., J.B. Yahaya, and I.B. Mohamed, *A model for enterprise resource planning implementation in the Saudi public sector organizations.* Heliyon, 2024. **10**(2).
8. Haddara, M. and A. Elragal, *ERP adoption cost factors identification and classification: a study in SMEs.* International Journal of Information Systems and Project Management, 2013. **1**(2): p. 5–21.
9. Lamoureux, S.M., H. Movassaghi, and N. Kasiri, *The role of government support in SMEs' adoption of sustainability.* IEEE Engineering Management Review, 2019. **47**(1): p. 110–114.
10. Santos, H., G. Lannelongue, and J. Gonzalez-Benito, *Integrating green practices into operational performance: Evidence from Brazilian manufacturers.* Sustainability, 2019. **11**(10): p. 2956.
11. Anjaria, K., *Enhancing sustainability integration in Sustainable Enterprise Resource Planning (S-ERP) system: Application of Transaction Cost Theory and case study analysis.* International Journal of Information Management Data Insights, 2024. **4**(2): p. 100243.
12. Ahn, B. and H. Ahn, *Factors affecting intention to adopt cloud-based ERP from a comprehensive approach.* Sustainability, 2020. **12**(16): p. 6426.
13. Anaya, L., L. Flak, and A. Abushakra, *Realizing Sustainable Value from ERP Systems Implementation.* Sustainability, 2023. **15**(7): p. 5783.
14. Santoso, R.W., et al., *Assessing the benefit of adopting ERP technology and practicing green supply chain management toward operational performance: An evidence from Indonesia.* Sustainability, 2022. **14**(9): p. 4944.
15. Adenekan, O.A., et al., *Enhancing manufacturing productivity: A review of AI-Driven supply chain management optimization and ERP systems integration.* International Journal of Management & Entrepreneurship Research, 2024. **6**(5): p. 1607–1624.
16. Tang, Y.M., et al., *Data-intensive inventory forecasting with artificial intelligence models for cross-border e-commerce service automation.* Applied Sciences, 2023. **13**(5): p. 3051.

17. Kitsantas, T., *Exploring Blockchain Technology and Enterprise Resource Planning System: Business and Technical Aspects, Current Problems, and Future Perspectives.* Sustainability, 2022. **14**(13): p. 7633.
18. Nofal, M.I., et al. *A Framework for Using Blockchain-Enabled Supply Chain Management to Enhance Transparency, Traceability, and Trust*. in *2024 2nd International Conference on Cyber Resilience (ICCR).* 2024. IEEE.
19. Mukred, M., et al. *Big data and electronic records management systems for public sector organizations competency*. in *2022 International Conference on Cyber Resilience (ICCR).* 2022. IEEE.
20. Abobakr, M.A., M. Abdel-Kader, and A.F. F. Elbayoumi, *An experimental investigation of the impact of sustainable ERP systems implementation on sustainability performance.* Journal of Financial Reporting and Accounting, 2024.
21. Anastasia, P., et al., *Assessing the role of green supply chain management on operational performance: mediating role of information technology infrastructure, internal and external integration Indonesian manufacturing.* Uncertain Supply Chain Management, 2024. **12**(4): p. 2227–2244.
22. Gimenez, C., V. Sierra, and J. Rodon, *Sustainable operations: Their impact on the triple bottom line.* International journal of production economics, 2012. **140**(1): p. 149–159.
23. Rasanjali, W., et al., *Implementing enterprise resource planning for lean waste minimisation: challenges and proposed strategies.* Smart and Sustainable Built Environment, 2024. **13**(2): p. 330–353.
24. Pérez Estébanez, R., *An Approach to Sustainable Enterprise Resource Planning System Implementation in Small-and Medium-Sized Enterprises.* Administrative Sciences, 2024. **14**(5): p. 91.
25. Ahmad, M.M. and R.P. Cuenca, *Critical success factors for ERP implementation in SMEs. Robotics and computer-integrated manufacturing*, 2013. **29**(3), p. 104–111.

Determinants of Biometric as a Service (BaaS) Adoption in the Telecommunication Companies: The Mediating Effect of Perceived Trust

Ahmad Abed Al-Hayy Al-Dalaien and Hairoladenan Kasim

Abstract The purpose of this research is to evaluate the function that Perceived Trust plays as a mediator in the adoption of biometric as a service (BaaS) among customers of Jordanian telecommunications. Through the utilization of a quantitative survey that was administered to four hundred employees, an investigation was carried out to investigate the ways in which organizational elements (Organizational Development and Training and Skill Development) and technological aspects (Compatibility, Complexity, and Trialability) influence BaaS User Acceptance. As a result of the data that we acquired, the significance of Corporate Culture, Training, and Skill Development was brought to light. According to the findings of a study, User Acceptance may be increased via the implementation of business culture and appropriate training. Compatibility emerged as a significant technical challenge that had an influence on adoption, underlining the significance of ensuring that there is a smooth interaction with those that are already in place. The links between organizational and technological characteristics and User Acceptance were also shown to be mediated by Perceived Trust in BaaS, a crucial mediating variable. Surprisingly, there was no significant effect of trialability on either acceptance or trust. Within the specific context of Jordanian telecommunications firms, our findings contribute to a better understanding of the complex interplay between organizational and technological elements that plays a role in molding workers' impressions of BaaS. This study highlights the importance of trust-building methods in the effective integration of this technology within the Jordanian telecommunications sector. It also proposes practical implications for companies who are working to increase the use of BaaS. In addition, it also highlights the determinants of BaaS adoption by the Jordanian telecommunications firms.

A. A. A.-H. Al-Dalaien (✉)
Cyber Security Department, Faculty of Science and Information, Jerash University, Jerash, Jordan
e-mail: a.aldalaien@jpu.edu.jo

H. Kasim
Universiti Tenaga Nasional (UNITEN), Jalan Kajang—Puchong, 43000 Kajang, Selangor, Malaysia
e-mail: Hairol@uniten.edu.my

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_22

Keywords Organizational culture · Training and skill development · Compatibility · Complexity · Trialability · Perceived trust in BaaS · User acceptance of BaaS · Jordan

1 Introduction

Biometrics-as-a-Service, often known as BaaS, is a paradigm that is hosted on the cloud and gives businesses access to biometric authentication services without requiring them to have substantial infrastructure or in-house knowledge [1]. In view of the fact that cyberattacks are inescapable and the security of sensitive information is a main issue, it is of the highest significance for enterprises working in the telecommunications industry to adopt Biometric as a Service for their operations. Biometric as a Service (BaaS) is a cutting-edge technology that has the ability to help enterprises improve their level of security. Through the use of biometric authentication, a comprehensive solution for cyber security may be obtained [2]. Because to the fact that they are responsible for managing sensitive consumer data, financial transactions, and key infrastructure, telecommunications businesses are in this position. Iris detection, face scanning, and fingerprint identification are all examples of biometric techniques that do not pose any dangers to the individual who uses them. Because of this, there is a decrease in both fraudulent activity and illegal access. BaaS contributes to the construction of a productive and secure working environment by enhancing cyber defenses and protecting sensitive information. This is accomplished via the protection of sensitive information. As a consequence of this, the usage of digital technology is becoming more widespread in the subject of telecommunications [3].

Cyber security has become crucial for enterprises, especially in Jordan, as we move farther into the digital era [4]. Organizations are increasingly exposed to a multitude of threats that might put their operations, reputation, and efficiency at risk as a result of their growing dependence on information and communication technology (ICT) to stimulate efficiency, innovation, and growth [5]. They risk reputations with cyberthreats. In the context of Jordanian enterprises, comprehensive cyber security standards implementation is critical. Cybersecurity also involves using plans, strategies, tactics, engagements, training, best practices, assurance, and expertise to defend information systems, organizations, and assets against cyberattacks. Information security is another name for cybersecurity [6].

The purpose of this study is to investigate the ways in which organizational and technological factors influence the Biometric as a Service (BaaS) variable Acceptance in Jordanian telecommunications companies, with a particular emphasis on the perspectives of workers [7]. The study was conducted in Jordan. By conducting this survey, we want to get an understanding of the reasons why professionals in the field of telecommunications use BaaS. Considering the culture of the organization, the training of its employees, and the compatibility, complexity, and trialability of the technology are all included in this. In order to better comprehend the intricate aspects that influence customer acceptability in the Jordanian telecommunications market, the study will also investigate Perceived Trust's Mediating Role in BaaS. The use of biometric technology is one of the topics covered in this article. This objective will be accomplished via the use of questionnaires and surveys for quantitative research [8].

2 Literature Review

2.1 Organizational Culture

The set of beliefs, traditions, and behavioral conventions that are disseminated and shared among the members of an organization is referred to as the organizational culture [9]. It is crucial to have a thorough awareness of the patterns in technology adoption in order to comprehend the connections between Organizational Culture and User Acceptance of BaaS in Jordanian Telecommunications Companies. This understanding is necessary in order to comprehend the linkages. In the event if the business puts a high value on adaptability, security, and innovation, it is possible that BaaS will be seen as an enhancement rather than a disruption [10, 11], and so on. In order to increase the number of users who are pleased with the BaaS, it is necessary to establish a culture that puts a high value on both technology and security. Consequently, the following hypothesis might be put forth:

H1 Organizational Culture has a significant effect on User Acceptance of BaaS in the Telecommunication Companies in Jordan.

2.2 Training and Skill Development

Training is described as the process of instructing an employee in a certain skill or kind of conduct in order to better fit in with the work [12]. Skill development is the process of recognizing, practicing, and internalizing skills in order to bring about improvements in the performance of abilities [13]. By evaluating the workers of Jordanian telecommunications businesses in terms of their Training and Skill Development as well as their User Acceptance of BaaS (Biometric as a Service), it is possible to ascertain whether or not the employees are ready to make use of cutting-edge technology. For the purpose of facilitating user adoption, staff people are required to grasp, implement, and troubleshoot BaaS systems for the organization. Training programs that are successful make these abilities and pieces of information accessible to participants [14]. Organizations provide their employees the ability to explore and make use of biometric technologies with confidence by giving them with complete training on the functions and security elements of BaaS. When employees believe they are appropriately prepared and competent in adopting BaaS into their everyday duties, there is a good association between Training and Skill Development initiatives and user adoption. Therefore, making investments in comprehensive training programs is not only a necessary step in the process of successfully adopting BaaS, but it is also a catalyst for the development of a workforce that is responsive to and accepting of this cutting-edge technology solution [15]. Therefore, one could hypothesize the following:

H2 Training and Skill Development has a significant effect on User Acceptance of BaaS in the Telecommunication Companies in Jordan.

2.3 Compatibility

Compatibility refers to the extent to which an invention is regarded to be compatible with the values that are now held, the values that have been held in the past, and the requirements of future adopters [16]. In order to understand how effectively this technology interfaces with the systems that are already in place inside Jordanian telecommunication firms, it is essential to grasp the link between Compatibility and User Acceptance of BaaS (biometric as a service) [17]. Compatibility is a measurement that determines how well BaaS aligns with the technical infrastructure that already exists, and its relevance rests in the fact that it helps to minimize interruptions to processes that have already been created. Employees are more likely to view BaaS as a helpful and integrated tool rather than a disruptive addition when it is able to fit in seamlessly with the structure that is already in place [11, 18]. When users perceive that BaaS has the ability to improve their day-to-day operations without imposing large modifications or inefficiencies, they develop favorable views about BaaS. This is because a high degree of compatibility creates positive attitudes. In summary, a positive association between Compatibility and User Acceptance highlights the significance of a harmonic integration between BaaS and the current technical environment in order to ensure a seamless and effective adoption within Jordan's telecommunication industry [19]. Therefore, one could hypothesize the following:

H3 Compatibility has a significant effect on User Acceptance of BaaS in the Telecommunication Companies in Jordan.

2.4 Complexity

Complexity is the amount and variety of components that make up an organization's internal and external environments [20]. When it comes to establishing whether or not employees in Jordanian telecommunication firms are prepared to accept this cutting-edge authentication technology, the link between Complexity and User Acceptance of BaaS (Biometric as a Service) is crucial. In the context of this discussion, the term "complexity" refers to the challenges and complexities that are thought to be connected with the implementation and use of BaaS. It is possible that employees may have increased degrees of fear and reluctance to incorporate this technology into their everyday duties as the difficulty of deploying BaaS rises [21]. A decrease in the adoption rate of the BaaS among consumers occurs when customers have the impression that the BaaS is extremely hard to understand. In order to increase the number of BaaS customers and decrease the perceived level of complexity associated with the telecommunications industry, it is necessary to have user-friendly interfaces, instructions that are easy to understand, and constant help. In order to properly incorporate biometric technology, it is vital to have an environment that is user-friendly, and this provides such environment [22]. Therefore, one could hypothesize the following:

H4 Complexity has a significant effect on User Acceptance of BaaS in the Telecommunication Companies in Jordan.

2.5 Trialability

Trialability is the degree to which an innovation may be experimented with on a restricted basis [20]. Trialability and User Acceptance of BaaS are crucial factors for Jordanian telecom carriers. The simplicity with which clients may try out a BaaS before choosing whether or not to use it is referred to as trialability in this context [23]. Employers may learn about the capabilities, benefits, and downsides of BaaS by testing it in a controlled environment. This is one method that potential employers can learn about BaaS. Through trial experiences, consumers are able to boost their degree of self-assurance and familiarity with technology, which eventually results in an improvement in the level of technology and user acceptance [24]. The implementation of this hands-on approach helps to cultivate a sense of user empowerment and alleviate anxieties, which in turn contributes to a more positive attitude and an improved readiness to adopt BaaS as a viable and user-friendly solution within the telecommunications industry. In essence, putting an emphasis on Trialability becomes a strategic strategy to encouraging user acceptability. This is accomplished by demystifying the technology and making the transition to its adoption less traumatic [25]. Therefore, one could hypothesize the following:

H5 Trialability has a significant effect on User Acceptance of BaaS in the Telecommunication Companies in Jordan.

2.6 Perceived Trust in BaaS

Perceived Trust is the amount of confidence that an individual has in another entity to carry out actions that are required of them without taking advantage of them [26]. User Acceptance of BaaS is highly impacted by Organizational Culture among Jordanian telecommunications businesses, with Perceived Trust serving as a mediating component in this relationship. An environment that encourages employees to create trust in the BaaS system is one that is fostered by a good corporate culture that is defined by openness to innovation and security consciousness. Because of the beneficial effect that organizational culture has on Perceived Trust, employees are more willing to embrace and incorporate BaaS into their routines. They view it as a secure and helpful addition to their work processes, which increases the likelihood that they will do so [27].

Effective Training and Skill Development programs contribute to Perceived Trust in BaaS, which in turn mediates the link between training activities and user acceptance. Employees get a heightened sense of competence and confidence in their ability to use the technology in a secure manner when they receive full training on the complexities of BaaS. Perceived Trust and BaaS Acceptability are positively connected with confidence levels. Training staff makes BaaS seem more trustworthy, which in turn leads to a rise in the number of individuals who utilize the platform [14].

Perceived Trust in BaaS and User Acceptance in Technology mediate the relationship between BaaS and Technology. workers are more likely to have trust in BaaS when the integration of the service does not result in a significant interruption to their information technology. This is because workers are expecting the service to be implemented. The confidence that people have in BaaS helps to boost both its adoption and its dependability. The implementation of BaaS will take place inside the framework of the firm if the workers think that it will complement the ways that they are presently using. There is a correlation between having a high degree of interoperability and an increased possibility that workers will trust BaaS to improve their workplace operations [28]. Hence, the mediating hypotheses are as follows:

- H6 Perceived Trust in BaaS significantly mediates the relationship between Organizational Culture and User Acceptance of BaaS in the Telecommunication Companies in Jordan.
- H7 Perceived Trust in BaaS significantly mediates the relationship between Training and Skill Development and User Acceptance of BaaS in the Telecommunication Companies in Jordan.
- H8 Perceived Trust in BaaS significantly mediates the relationship between Compatibility and User Acceptance of BaaS in the Telecommunication Companies in Jordan.
- H9 Perceived Trust in BaaS significantly mediates the relationship between Complexity and User Acceptance of BaaS in the Telecommunication Companies in Jordan.
- H10 Perceived Trust in BaaS significantly mediates the relationship between Trialability and User Acceptance of BaaS in the Telecommunication Companies in Jordan.

2.7 Adoption of Biometric as a Service (BaaS)

This research makes use of both the Social Cognitive Theory and the Diffusion of Innovation Theory in order to analyze the adoption of BaaS by Jordanian telecoms providers [29]. The purpose of this research, which is grounded on Social Cognitive Theory, is to analyze the ways in which organizational elements like Organizational Culture and Training & Skill Development impact the perceptions and behaviors of workers in relation to BaaS. Additionally, the Innovation Theory is being used in order to explore the ways in which technical qualities such as Compatibility, Complexity, and Trialability have an impact on the dissemination of BaaS. There is a connection between the effects of organizational and technological elements and the eventual user acceptability, and one of the variables that has been identified as a mediating variable is perceived trust in BaaS. This integrated theoretical framework offers a complete lens for understanding the complex dynamics of BaaS adoption. It takes into consideration the interaction between organizational and technical factors within the context of Jordanian telecommunication businesses, which is a singular setting (see Fig. 1).

3 Instrument Development

Within the context of this study, the creation of instruments was carried out with great attention in order to accurately reflect nature. In light of this, the questionnaire was developed to contain a total of 35 items, and the variables were evaluated using a Likert scale with five points, where five represents "Strongly Agree" and one represents "Strongly Disagree" [30]. In addition, the validated instruments that are

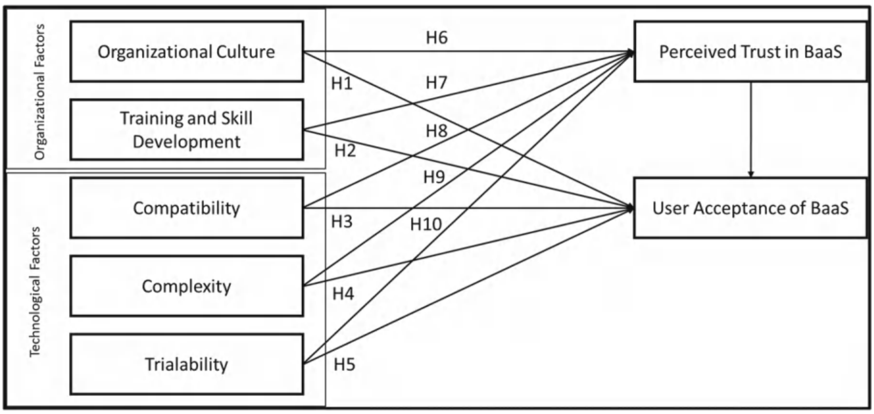


Fig. 1 Research framework

Table 1 Questionnaire development

Variable	No. of items	References
Organizational culture	5	[10, 31]
Training and skill development	5	[14, 15]
Compatibility	5	[17, 18, 32]
Complexity	5	[21, 22]
Trialability	5	[23, 25]
Perceived trust in BaaS	5	[28, 33]
User acceptance of BaaS	5	[27, 34]

presented in Table 1 were taken from similar earlier research in order to measure the variables that were being investigated in this investigation.

4 Research Methodology

For the purpose of this study, a quantitative research technique was utilized to evaluate the characteristics that influence employee adoption of Biometric as a Service (BaaS) in the mobile service industry in Jordan. For the purpose of data collection, a questionnaire that was self-administered and delivered to 400 people working within the industry was utilized. In order to guarantee completeness, clarity, and compliance with the standards, the questionnaire was subjected to preliminary testing. The validity and reliability of the questionnaire were so ensured.

A cutting-edge tool called SmartPLS 4, which is a Partial Least Squares Structural Equation Modeling application, was used in order to process the data. Because it was able to handle non-normal data, small sample sizes, and intricate models with mediating effects, this methodology was perfect for use in scientific research. For the purpose of this study, the SmartPLS 4 measuring model was used, and it underwent a comprehensive evaluation to ensure its construct validity and reliability. It was via the use of the variable model that the hypothesized correlations between the variables were successfully attained. This came after the step that came before it.

5 Data Analysis

5.1 Measurement Model

The measurement model is a sub-model in structural equation modeling that specifies the indicators for each construct and assesses the reliability of each construct for estimating the causal relationships. With the help of Smart PLS 4, the research model for this study was put through its paces. In addition, an investigation was carried out with reference to "the measurement model (validity and reliability of the measures)" and "the structural model (testing the hypothesized linkages within the framework)". According to the recommendations made by [35], these values are more than the predetermined threshold for factor loadings. As a result, no modification considered on the research model. According to the data shown in Table 2 and Fig. 2, all of the variables have, in general, reached the threshold for the factor loadings, Cronbach Alpha, composite reliability, and average variance extracted (AVE) values.

An assessment was carried out in the second phase of the process to assess the discriminant validity. This was done to determine the precise extent to which a certain construct differs from other constructs [36, 37]. As indicated by [37], the estimate of the model did not surpass 0.95. Measures of "the correlations between constructs and the square root of the average variance" found for a construct were used to evaluate the validity of the model [36, 37]. Accordingly, Table 3 presents the findings of the Fornell and Larcker Criterion and indicates that there is no value greater than the suggested cutoff point of 0.95 [36].

Furthermore, in the case of perfect measurement, or complete reliability, the Heterotrait-Monotrait ratio (HTMT) approximates the true correlation between two constructs. Furthermore, the average of all correlations between indicators measuring different constructs with respect to the (geometric) mean of the average correlations between indicators measuring the same construct is represented by the Heterotrait-Monotrait correlations (HTMT). This measure can be employed for the evaluation of discriminant validity [35]. The literature recommends a level of HTMT of 0.90 [38] (see Table 4).

5.2 *Structural Model*

Structural model shows the causal and correlational links among latent variables in a theoretical model. The theoretical or conceptual component of the path model is represented by the structural model. The structural model in PLS-SEM, sometimes called the inner model, comprises the latent variables and their path relationships [35]. After the measurement model has been evaluated, the structural model has to be evaluated. According to [35], evaluating the structural model in accordance with PLS-SEM requires four steps: evaluating collinearity (step one), evaluating path coefficients (step 2), determining the coefficient of determination (R^2 value) (step 3), and determining the effect size (f^2) [35].

Table 5 illustrates the results of PLS bootstrapping consisting of "the Beta value, t-values, p-values, hypothesis results (whether supported or not) confidence interval, f^2, and VIF scores". Furthermore, Fig. 3 summarizes the results of the structural model and PLS bootstrapping.

Table 2 Convergent validity

Items	Outer loadings	Cronbach's alpha	Composite reliability	Average variance extracted (AVE)
ACC1	0.754	0.835	0.883	0.603
ACC2	0.729			
ACC3	0.785			
ACC4	0.831			
ACC5	0.778			
COM1	0.765	0.873	0.908	0.664
COM2	0.848			
COM3	0.819			
COM4	0.850			
COM5	0.789			
COX1	0.826	0.924	0.943	0.766
COX2	0.887			
COX3	0.890			
COX4	0.871			
COX5	0.900			
OC1	0.774	0.885	0.916	0.686
OC2	0.882			
OC3	0.876			
OC4	0.819			
OC5	0.783			
PT1	0.695	0.813	0.870	0.573
PT2	0.749			
PT3	0.808			
PT4	0.808			
PT5	0.718			
TRA1	0.723	0.796	0.853	0.538
TRA2	0.648			
TRA3	0.823			
TRA4	0.733			
TRA5	0.730			
TSD1	0.729	0.867	0.903	0.652
TSD2	0.821			
TSD3	0.830			
TSD4	0.806			
TSD5	0.847			

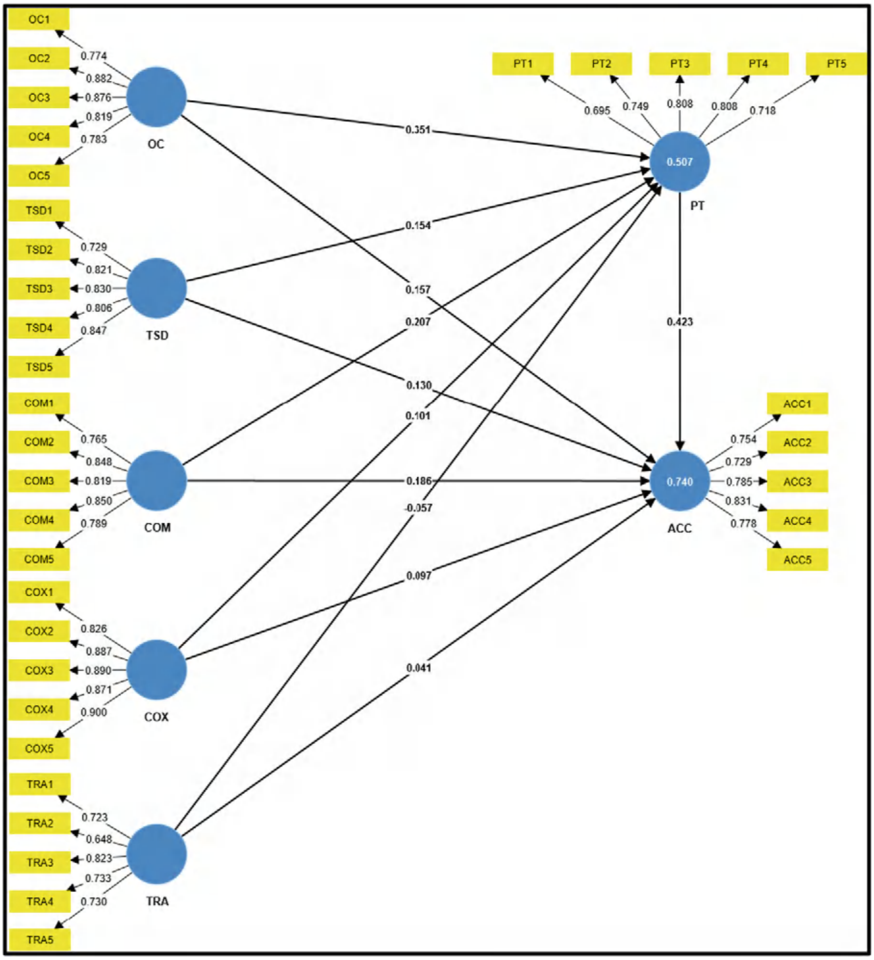


Fig. 2 PLS algorithm results

Table 3 Discriminant validity—Fornell and Larcker criterion

	ACC	COM	COX	OC	PT	TRA	TSD
ACC	0.776						
COM	0.722	0.815					
COX	0.641	0.639	0.875				
OC	0.720	0.667	0.625	0.828			
PT	0.779	0.616	0.551	0.655	0.757		
TRA	0.154	0.169	0.121	0.154	0.065	0.734	
TSD	0.726	0.777	0.678	0.724	0.630	0.132	0.808

Table 4 Discriminant validity—HTMT

	ACC	COM	COX	OC	PT	TRA	TSD
ACC							
COM	0.841						
COX	0.731	0.710					
OC	0.834	0.759	0.690				
PT	0.827	0.726	0.634	0.764			
TRA	0.170	0.178	0.133	0.177	0.096		
TSD	0.840	0.881	0.755	0.828	0.733	0.144	

Table 5 PLS bootstrapping results

H	Path	Beta	Std error	t-value	p-value	f^2	VIF	R^2
H1	OC → ACC	0.157	0.049	3.191	0.001	0.356	2.603	0.740
H2	TSD → ACC	0.130	0.058	2.239	0.013	0.219	2.410	
H3	COM → ACC	0.186	0.053	3.493	0.000	0.446	2.885	
H4	COX → ACC	0.097	0.038	2.532	0.006	0.218	2.087	
H5	TRA → ACC	0.041	0.026	1.569	0.058	0.006	1.040	
H6	OC → PT → ACC	0.149	0.030	5.035	0.000	0.406	2.353	
H7	TSD → PT → ACC	0.065	0.027	2.388	0.008	0.214	2.362	
H8	COM → PT → ACC	0.088	0.026	3.354	0.000	0.381	2.798	
H9	COX → PT → ACC	0.043	0.021	2.070	0.019	0.210	2.066	
H10	TRA → PT → ACC	0.024	0.016	1.472	0.071	0.002	1.034	

Assessment of the structural model for collinearity issues

The structural model's assessment of collinearity problems comes first. Preventing collinearity between the constructs is crucial in order to carry out a latent variable analysis in the structural model. Therefore, by measuring the VIF value, the collinearity has been determined. In accordance with [39] proposal, the assessment's threshold value is 3.3 [39]. In this study, as illustrated in Table 5, for the constructions, every inner VIF value falls between 1.034 and 2.885. Since they are all smaller than 3.3, collinearity is not an issue for this research.

Assessing the significance of the structural model relationships

As shown in Table 5, the bootstrapping approach has been used to generate information for every path relationship in the model in order to assess the hypotheses.

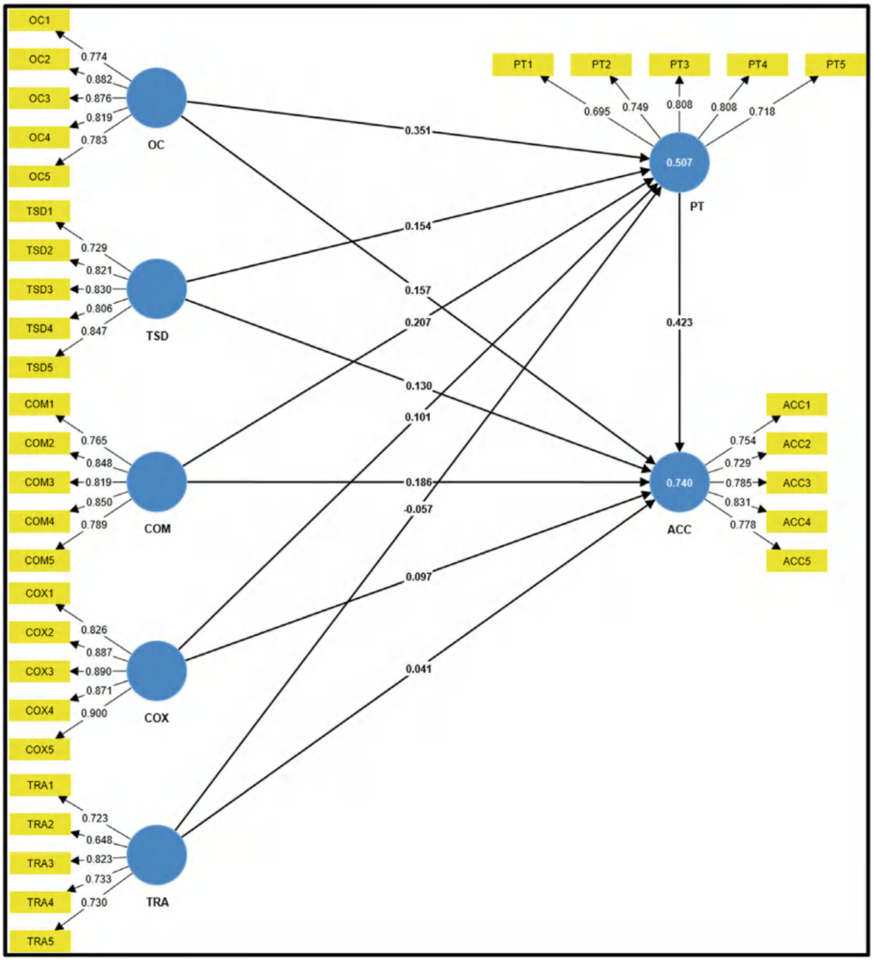


Fig. 3 PLS bootstrapping results

As shown in Table 5, the bootstrapping approach has been used to generate data for every path relationship in the model in order to test the hypotheses.

To create a bootstrap sample and provide standard errors for hypothesis testing, PLS bootstrapping is a nonparametric test that consists of repeated random sampling with replacement from the original sample [35]. In addition, [40] recommended using 1000 samples for bootstrapping for determining the number of resampling [40]. For the constructs in this research, ten hypotheses have been established. Using SmartPLS 4's bootstrapping tool, t-statistics for each path have been produced in order to test the significance level. A one-tailed test, 1000 subsamples, and a significance level of 0.05 have been specified for the bootstrapping. For the one-tailed test, the critical value at the five percent significance level ($\alpha = 0.05$) is 1.645 [41].

The value of the path coefficients has a standardized value that is about between −1 and + 1 (values from 0.027 to 0.321) based on the data shown in Table 5. Path coefficients with an estimated value near to + 1 indicate strong positive relationships, whereas values closer to 0 indicate weaker relationships [35]. The t-test is then performed on the relationships, and it is discovered that the t-values are greater than or equal to 1.645. In light of this, the relationships for H1, H2, H3, H4, H6, H7, H8, and H9 are significant at 0.05. Refer to Table 5.

The Coefficient of Determination (R^2)

The calculated coefficient of determination (R^2) value is then used to assess the model's prediction accuracy. A greater value indicates a better level of predicted accuracy. R^2 is a measure of the predictive capacity of the model and runs from zero to one [35]. The value of R^2 has been computed using the Smart PLS technique, as seen in Table 5.

This research has adhered to the recommendations established by [42], identifying values of "0.02, 0.13, and 0.26" as representing a "weak, moderate, and substantial level of predictive accuracy", in light of the fact that there are several sets of regulations pertaining to the accepted value of R^2 [42]. Overall, referring to Table 5, Organizational Culture, Training and Skill Development, Compatibility, Complexity, Trialability, and Perceived Trust in BaaS explain 74 percent of the variance in User Acceptance of BaaS. This signifies a substantial level of predictive accuracy.

Assessment of the effect size (f^2)

The effect sizes (f^2) have been assessed at this point. The proportionate influence of a predictor construct on endogenous constructs is related to the value of f^2. Moreover, [43] state that in addition to the p-value, it is essential to disclose the substantive significance (effect size) as well as the statistical significance (p-value) [42]. Moreover, a [44] guideline has been adhered to in order to calculate the effect size [43]. Moreover, [44] research yielded values of 0.02, 0.15, and 0.35, which correspond to weak, medium, and substantial effects, respectively [44]. As it can be viewed in Table 5, The production of the value of R^2 for User Acceptance of BaaS is somewhat affected by Training and Skill Development and Complexity, while the production of the value of R2 for User Acceptance of BaaS is significantly affected by Organizational Culture and Compatibility. Trialability, however, has no impact on the value of R^2 for User Acceptance of Production.

6 Discussion

The purpose of this study is to evaluate the factors that influence the User Acceptance of Biometric as a Service (BaaS) among personnel working in the telecom industry in Jordan. This research investigates the variable Acceptance of BaaS in connection to the effects of Organizational Variables (such as Organizational Culture and Training and Skill Development) and Technological Variables (such as such as

Compatibility, Complexity, and Trialability). Specifically, the study focuses on the relationship between these two types of variables. In particular, a questionnaire and a quantitative research are used in order to effectively accomplish this goal. This study also looks into Perceived Trust as a mediating variable. The objective of this study is to analyze the ways in which variables related to organizations and technologies, specifically those linked with the Social Cognitive Theory and the Diffusion of Innovation Theory, have an impact on the manner in which users embrace BaaS. An investigation of the advances that have occurred in the implementation of BaaS by Jordanian telecommunications firms is presented in this study.

It is crucial to recognize that the prospects for Jordanian telecoms customers to adopt biometric as a service are significantly affected by the culture of the firm. It would seem that a business culture that puts a high value on innovation, security, and technological improvements has a significant effect on the positive sentiments that employees have on Biometric as a Service (BaaS) and their readiness to adopt it. Companies that put a significant emphasis on flexibility and understand the significance of cutting-edge technology in enhancing operational efficiency are more likely to see the adoption of BaaS by workers as a secure and useful addition to their work processes. This is because BaaS is a secure and beneficial addition to the processes that workers use in their job. This research highlights the strategic significance of a healthy corporate culture in order to support the development and acceptance of platform as a service (BaaS) inside Jordan's fast developing telecommunications ecosystem [31]. This aspect has been discussed previously in the literature review, some of them agreed with the results and others did not [8, 31].

Secondly, Training and Skill Development had a significant influence on Biometric as a Service (BaaS) User Acceptance in Jordanian telecoms firms. In order to expand the number of workers who accept the technology, it is essential to give them with information and skills that are linked to BaaS. It has been reported that users of BaaS who take part in training programs that are successful report feeling more at ease and educated, which in turn boosts the usability and security of the technology. It is thus necessary for Jordanian telecommunications firms to make investments in comprehensive training and skill development in order to increase user readiness and adoption of BaaS. Overall, many studies were conducted to highlight this relationship [12–14]

The third result demonstrates the significance of compatibility on User Acceptance of Biometric as a Service (BaaS) among Jordanian Telecommunication Companies. The extent to which a company's processes and systems are compatible with the software as a service (BaaS) is one of the factors that determines the adoption of technology. It is possible to achieve the goal of increasing the possibility that workers would see BaaS as a valued tool, which will in turn raise adoption, by integrating it in a seamless way into the existing technological infrastructure. In light of this, the strategic significance of merging BaaS with the technology that is now accessible is brought into focus. In the fast increasing field of telecommunications in Jordan, compatibility is also a crucial component in the process of building an environment that is suitable to BaaS [17, 45].

The fourth significant finding emphasizes the significant impact that complexity has on the BaaS User Acceptance of Jordanian Telecommunication Companies. There is a significant effect on the degree to which workers embrace BaaS technology. This is because of the perceived difficulty and complexity of the technology. There is a risk that the complexity of BaaS might further enhance the level of uneasiness and reluctance that employees have against its installation. Therefore, it is vital to have user-friendly interfaces, instructions that are simple to comprehend, and ongoing help in order to reduce the perceived complexity of biometric as a service (BaaS) and to increase user adoption within the organizational structures of Jordanian mobile telecommunications companies [21].

The sixth significant discovery is that the trialability of Biometric as a Service (BaaS) does not have any influence on the User Acceptance of the service by Jordanian telecommunications companies. It would seem, on the basis of the information that has been provided, that the provision of workers with the option to utilize BaaS in a controlled setting does not have any effect on the pace at which they adopt the technology. Workers may not perceive significant benefits or advantages that significantly impact BaaS adoption, despite the fact that they have evaluated trial versions of the software. This may be the case even if they have been testing the program. This is shown by a factor that does not have a significant impact. Through the course of this research, it is shown that the complexity of BaaS acceptance variables has grown over the course of time. These results imply that variables other than trial experiences alone have a significant influence on how Jordanian telecommunications company workers view biometric technology [25].

Perceived Trust in BaaS is a significant mediator of the relationship between Organizational Culture and User Acceptance of BaaS in Jordanian Telecommunications Businesses, as shown by the sixth revealing result, which underlines the significance of this mediator. This illustrates that the effect of a good company culture on user acceptance is channeled via BaaS confidence. When workers perceive that their company culture fosters innovation and security, they are more inclined to trust BaaS. After the company has implemented the technology, this leads to an improvement in attitudes and acceptance of the technology at the organization. Furthermore, the significance of trust as a mediating component in the process of translating the cultural values of organizations into BaaS is emphasized by this study.

According to the sixth significant result, Perceived Trust in BaaS has a significant role in mediating the relationship between Training and Skill Development and User Acceptance of BaaS among Jordanian telecommunications firms. This finding was made possible by the fact that the perceived trust in BaaS was found to be significant. This demonstrates that the faith that employees have in the capabilities and security of BaaS as a consequence of these training activities is a significant role in the positive effect that successful training programs have on user adoption. Also, this demonstrates that the training efforts have been successful. It is because of the training efforts that have been undertaken that this faith has been developed. Consequently, this transpired as a consequence of the execution of these trainings. Workers' trust in BaaS is increased when they have access to more information and better skills, which in turn leads to a rise in the use of technology and a more favorable attitude about it.

For BaaS to be used more widely, training and trust are crucial elements. This study came to the conclusion that both of the components contribute significantly to the overall picture [34].

Perceived Trust in BaaS is a significant mediator of the relationship between User Acceptance and Compatibility in Jordanian telecommunications firms, according to the eighth main result. This discovery was made by the researchers. The confidence that users have in technology seems to be more essential than compatibility when it comes to increasing user adoption. When a BaaS is compatible with their System, employees are more likely to adopt it. As a consequence of this, they have a greater expectation that the integration would be carried out without any interruptions. When it comes to the compatibility of BaaS, trust is a significant feature that functions as a mediating component among the many factors that are present in Jordan [33].

The tenth significant conclusion is that the Perceived Trust in BaaS does not significantly moderate the connection between Trialability and User Acceptance across Jordanian telecommunication providers. The provision of workers with the opportunity to utilize BaaS has an influence on the manner in which they engage with technology; nevertheless, it does not have an effect on trust as a mediator for user acceptance. The lack of mediation in this relationship is evidence that the trialability of Biometrics as a service (BaaS) could not have a direct influence on the level of trust in the particular service. It is very evident that Jordanian telecommunications businesses need a more sophisticated knowledge of trust and trialability in light of the aforementioned circumstances.

7 Future Research Recommendation

A number of prospective research subjects have been uncovered as a result of the outcomes of this study. These themes will be of great assistance to us in acquiring a better knowledge of Biometric as a Service (BaaS) User Acceptance in telecoms companies. In the future, it is feasible that research will be conducted to explore the characteristics of corporate culture that have the greatest influence on trust and user approval respectively. By first getting a grasp of the business culture, which includes characteristics such as leadership support, communication methods, and collaboration, it may be feasible to create an environment that is conducive to the use of BaaS.

The comparison of Training and Skill Development programs is something that should be proposed for future research in order to determine which approaches are the most beneficial in terms of enhancing BaaS competence and trust. For instance, investigating the ways in which interactive training modules, simulations of real-world events, and continuous learning programs affect the attitudes of workers and the adoption of BaaS systems is something that may be done.

It is necessary to do further study on the aspects of compatibility that foster acceptance and trust in order to overcome the issues that are associated with technology. It is possible that creating the major compatibility requirements that develop customer confidence might be useful for telecom companies that are pursuing BaaS adoption plans. This is something that should be considered in the future studies.

There is a possibility that BaaS will be simplified and trusted in the future research endeavours. It may be advantageous to study the ways in which instructional materials, support systems, and user interface design have the potential to simplify the user experience and decrease complexity. This might be a valuable endeavor.

There is a possibility that future research may analyze the intricacies of trial experiences and how they impact the emotions of consumers. This is despite the fact that the present study did not identify any significant mediating effect on trust. There are a number of trialability variables that may reveal features that impact trust and acceptance. Some of these variables include trial duration, user feedback methods, and trial conditions.

References

1. R. Thanki, S. Borra, and A. Kothari, "Fragile Watermarking Framework for Tamper Detection of Color Biometric Images," *Int. J. Digit. Crime Forensics*, vol. 13, no. 2, pp. 35–56, 2021, https://doi.org/10.4018/IJDCF.2021030103.
2. P. S. Jangid and V. A. Bharadi, "Evolving Authentication Design Consideration and BaaS Architecture for Internet of Biometric things," in *2020 International Conference on Convergence to Digital World-Quo Vadis (ICCDW)*, 2020, pp. 1–7.
3. Y. Moolla, A. De Kock, G. Mabuza-Hocquet, C. S. Ntshangase, N. Nelufule, and P. Khanyile, "Biometric recognition of infants using fingerprint, iris, and ear biometrics," *IEEE Access*, vol. 9, no. 1, pp. 38269–38286, 2021.
4. A. A. S. Alrkeebat, "Cyber Security and Cyber Threats in the Jordan Business Context," *Int. J. Acad. Res. Bus. Soc. Sci.*, vol. 14, no. 6, pp. 1872–1883, 2024, https://doi.org/10.6007/IJARBSS/v14-i6/21991.
5. M. A. Al-ma'aitah, "Investigating the drivers of cybersecurity enhancement in public organizations: The case of Jordan," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 88, no. 5, p. e12223, Sep. 2022, https://doi.org/10.1002/isd2.12223.
6. T. B. Amer and M. I. A. Al-Omar, "The impact of cyber security on preventing and mitigating electronic crimes in the Jordanian banking sector," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, 2023, https://doi.org/10.14569/IJACSA.2023.0140841.
7. C. Wu, M. J. Amiri, J. Asch, H. Nagda, Q. Zhang, and B. T. Loo, "FlexChain: an elastic disaggregated blockchain," *Proc. VLDB Endow.*, vol. 16, no. 1, pp. 23–36, 2022.
8. K. Bozkus, *Organizational Culture Change and Technology: Navigating the Digital Transformation*. NY: IntechOpen, 2023.
9. A. Vecchi and V. Vallisi, "Supply Chain Resilience," in *Handbook of Research on Global Supply Chain Management*, B. Christiansen, Ed. Hershey, PA, USA: IGI Global, 2016, pp. 147–160. https://doi.org/10.4018/978-1-4666-9639-6.ch009.
10. O. Bello and O. Olanrewaju, "Factors influencing biometric technology adoption: Empirical evidence from Nigeria," *African J. Sci. Technol. Innov. Dev.*, vol. 14, no. 2, pp. 392–404, 2022.
11. F. H. Zawaideh, W. Abu-ulbeh, Y. I. Majdalawi, M. D. Zakaria, J. A. Jusoh, and S. Das, "E-Commerce Supply Chains with Considerations of Cyber-Security," in *2023 International Conference on Computer Science and Emerging Technologies (CSET)*, 2023, pp. 1–8.
12. K. Agha, "Training and Upgrading Skills of Employees Towards Building Resilience Among the Workforce," in *Handbook of Research on Supply Chain Resiliency, Efficiency, and Visibility in the Post-Pandemic Era*, Y. Ramakrishna, Ed. Hershey, PA, USA: IGI Global, 2022, pp. 226–240. https://doi.org/10.4018/978-1-7998-9506-0.ch012.

13. D. C. Hurst, R. Clapperton, R. J. Dixon, and M. T. Morpurgo, "Why a Virtual Co-Op?: A Case Study on an Artificially-Intelligent Work Integration Simulation," in *Developments in Virtual Learning Environments and the Global Workplace*, S. Swartz, B. Barbosa, I. Crawford, and S. Luck, Eds. Hershey, PA, USA: IGI Global, 2021, pp. 289–309. https://doi.org/10.4018/978-1-7998-7331-0.ch015.
14. M. Hawkins, "Virtual employee training and skill development, workplace technologies, and deep learning computer vision algorithms in the immersive metaverse environment," *Psychosociological Issues Hum. Resour. Manag.*, vol. 10, no. 1, pp. 106–120, 2022.
15. G. Liu-Farrer, B. S. Yeoh, and M. Baas, "Social construction of skill: An analytical approach toward the question of skill in cross-border labour mobilities," *Journal of Ethnic and Migration Studies*, vol. 47, no. 10. Taylor & Francis, pp. 2237–2251, 2021.
16. R. Ramanathan, L. W. L. Ko, H. Chen, and U. Ramanathan, "A Study on Green Characteristics of RFID using Innovation Diffusion Theory," in *Encyclopedia of E-Commerce Development, Implementation, and Management*, I. Lee, Ed. Hershey, PA, USA: IGI Global, 2016, pp. 1913–1922. https://doi.org/10.4018/978-1-4666-9787-4.ch135.
17. P. Lissin *et al.*, "Relative advantages and compatibility of a biometric patient identification tool in Zambia: a qualitative analysis," *Gates Open Res.*, vol. 5, no. 102, p. 102, 2021.
18. S. O. Olabanji, O. O. Olaniyi, C. S. Adigwe, O. J. Okunleye, and T. O. Oladoyinbo, "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," *Auth. Access Control within Cloud-Based Syst. (January 25, 2024)*, vol. 1, no. 1, pp. 1–10, 2024.
19. B. Singh and N. Singh, "MoLaBSS: Server-Specific Add-On Biometric Security Layer Model to Enhance the Usage of Biometrics," *Information*, vol. 11, no. 6, p. 308, 2020.
20. R. Korn, "Empathy First: Refurbishing a Teams' Approach to Student Success," in *Applying Design Thinking to the Measurement of Experiential Learning*, A. Peck and D. DeSawal, Eds. Hershey, PA, USA: IGI Global, 2021, pp. 12–30. https://doi.org/10.4018/978-1-7998-7768-4.ch002.
21. M. R. S. G. N. Sawant, "Permission Blockchain based Smart Contract utilizing Biometric Authentication as a Service." University of Mumbai.
22. A. Bodepudi and M. Reddy, "Spoofing attacks and mitigation strategies in biometrics-as-a-service systems," *Eig. Rev. Sci. Technol.*, vol. 4, no. 1, pp. 1–14, 2020.
23. M. A. Nooria and M. N. Jailanib, "Biometric authentication technology acceptance in mobile banking in Iraq." UKM, Bangi, 2022.
24. M. A. N. M. N. Jailani, "Biometric authentication technology acceptance in mobile banking in Iraq," *NeuroQuantology*, vol. 20, no. 7, p. 3004, 2022.
25. C. Kim, K. C. Lee, and F. J. Costello, "The intention of passengers towards repeat use of biometric security for sustainable airport management," *Sustainability*, vol. 12, no. 11, p. 4528, 2020.
26. C. Phonthanukitithaworn, C. Sellitto, and M. W. L. Fong, "A Study of Mobile Payment (M-Payment) Services Adoption in Thailand," in *Encyclopedia of Information Science and Technology, Third Edition*, D. B. A. Khosrow-Pour Mehdi, Ed. Hershey, PA, USA: IGI Global, 2015, pp. 731–741. https://doi.org/10.4018/978-1-4666-5888-2.ch070.
27. D. Liu and W. Tu, "Factors influencing consumers' adoptions of biometric recognition payment devices: combination of initial trust and UTAUT model," *Int. J. Mob. Commun.*, vol. 19, no. 3, pp. 345–363, 2021.
28. N. Singh and N. Sinha, "How perceived trust mediates merchant's intention to use a mobile wallet technology," *J. Retail. Consum. Serv.*, vol. 52, p. 101894, 2020.
29. W. R. Malatji, R. van Eck, and T. Zuva, "Acceptance of biometric authentication security technology on mobile devices," in *2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, 2020, pp. 1–5.
30. R. W. Brislin, "Back-Translation for Cross-Cultural Research," *J. Cross. Cult. Psychol.*, vol. 1, no. 3, pp. 185–216, Sep. 1970, https://doi.org/10.1177/135910457000100301.

31. F. Huang and T. Teo, "Influence of teacher-perceived organisational culture and school policy on Chinese teachers' intention to use technology: An extension of technology acceptance model," *Educ. Technol. Res. Dev.*, vol. 68, no. 3, pp. 1547–1567, 2020.
32. L. Norfolk and M. O'Regan, "Biometric technologies at music festivals: An extended technology acceptance model," in *Journal of Convention & Event Tourism*, 2020, vol. 22, no. 1, pp. 36–60.
33. R. R. Ilhamalimy and H. Ali, "Model perceived risk and trust: e-WOM and purchase intention (the role OF trust mediating IN online shopping IN shopee Indonesia)," *Dinasti Int. J. Digit. Bus. Manag.*, vol. 2, no. 2, pp. 204–221, 2021.
34. E. Moriuchi, "An empirical study of consumers' intention to use biometric facial recognition as a payment method," *Psychol. Mark.*, vol. 38, no. 10, pp. 1741–1765, 2021.
35. J. F. Hair, G. T. M. Hult, C. Ringle, and M. Sarstedt, *A primer on partial least squares structural equations modeling (PLS-SEM)*, 2nd ed. Los Angeles: SAGE, 2017.
36. C. Fornell and D. F. Larcker, "Evaluating Structural Equation Models with Unobservable Variables and Measurement Error," *J. Mark. Res.*, vol. 18, no. 1, pp. 39–50, 1981, https://doi.org/10.2307/3151312.
37. R. B. Kline, *Principles and practice of structural equation modeling*, 4th ed. New York: The Guilford Press, 2016.
38. A. H. Gold, A. Malhotra, and A. H. Segars, "Knowledge Management: An Organizational Capabilities Perspective," *J. Manag. Inf. Syst.*, vol. 18, no. 1, pp. 185–214, May 2001, https://doi.org/10.1080/07421222.2001.11045669.
39. A. Diamantopoulos and J. A. Siguaw, "Formative Versus Reflective Indicators in Organizational Measure Development: A Comparison and Empirical Illustration," *Br. J. Manag.*, vol. 17, no. 4, pp. 263–282, Dec. 2006, https://doi.org/10.1111/j.1467-8551.2006.00500.x.
40. W. W. Chin, "How to Write Up and Report PLS Analyses," in *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V. Esposito Vinzi, W. W. Chin, J. Henseler, and H. Wang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 655–690. https://doi.org/10.1007/978-3-540-32827-8_29.
41. T. Ramayah, J. Cheah, F. Chuah, H. Ting, and M. A. Memon, "Partial least squares structural equation modeling (PLS-SEM) using smartPLS 3.0," in *An Updated Guide and Practical Guide to Statistical Analysis*, Pearson., 2018.
42. A. Cohen, "Comparison of correlated correlations," *Stat. Med.*, vol. 8, no. 12, pp. 1485–1495, Dec. 1989, https://doi.org/10.1002/sim.4780081208.
43. G. Sullivan and R. Feinn, "Using Effect Size—or Why the P Value Is Not Enough," *J. Grad. Med. Educ.*, vol. 4, no. 1, pp. 279–282, Sep. 2012, https://doi.org/10.4300/JGME-D-12-00156.1.
44. J. Cohen, *Statistical power analysis for the behavioral sciences*, 2nd ed. Hillsdale, N.J.: L. Erlbaum Associates, 1988.
45. F. H. Zawaideh, W. A. H. M. Ghanem, M. H. Yusoff, S. I. A. Saany, J. A. Jusoh, and Y. A. B. El-Ebiary, "The Layers of Cloud Computing Infrastructure and Security Attacking Issues," *J. Pharm. Negat. Results*, pp. 792–800, 2022.

The State of the Art in AI-Based Phishing Detection: A Systematic Literature Review

Mohammed Alghenaim, Gamal Alkawsi, and Christopher R. Barnhart

Abstract Phishing is a prevalent social engineering attack that deceives users into revealing personal or sensitive data. The application of artificial intelligence (AI) techniques in detecting phishing attacks are explored in this literature review, highlighting machine learning (ML) and deep learning (DL) algorithms. The study aims to consolidate existing research, identify effective AI algorithms, and assess their performance. The study was conducted as a survey, with data gathered from IEEE Xplore, SpringerLink, MDPI, and ScienceDirect using PRISMA guidelines. The review included 21 peer-reviewed articles published between 2019 and 2024. The results show that AI-based methods, particularly hybrid models and convolutional neural networks (CNNs), have high accuracy in phishing detection, with XG Boost achieving the highest accuracy at 99.89%, followed by PILFER with 99.5%. Random Forest (RF) showed consistent performance across multiple studies. The review sheds light on the importance of diverse feature sets and numerous classifiers to enhance detection accuracy. However, challenges such as the evolving nature of phishing attacks and imbalanced datasets persist. The study recommends continuous updates to detection models and integrating multiple data sources for improved efficacy. Future research could usefully explore optimized learning parameters and hybrid approaches to enhance detection accuracy and scalability.

Keywords Classification · Phishing attacks · Systematic literature review · Artificial intelligence · Machine learning

M. Alghenaim (✉) · C. R. Barnhart
EC-Council University, 101 Sun Ave NE C, Albuquerque, NM 87109, USA
e-mail: Mohammed.alghenaim@eccu.edu

G. Alkawsi
Institute of Informatics and Computing in Energy (IICE), The Energy University, 43000 Kajang, Selangor, Malaysia

M. A. Al-Sharafi et al. (eds.), *Current and Future Trends on AI Applications*, Studies in Computational Intelligence 1178, https://doi.org/10.1007/978-3-031-75091-5_23

1 Introduction

Phishing is a cyberattack that utilizes emails, text messages, websites, or phone calls to deceive individuals into revealing sensitive data [1]. It is a social engineering attack commonly employed by cybercriminals to gain unauthorized access to personal and sensitive data such as credit card data, usernames, and passwords [2–4]. Typically, phishing involves the fraudulent acquisition of sensitive data by marauding reputable entities to deceive individuals into revealing such information through social engineering [5]. Phishing methods are diverse and leverage various communication channels such as email, social media platforms, or instant messaging [6]. Hackers use vulnerable internet systems and networks to target computers or mobile devices. Similarly, phishing emails are usually compromised to defraud individuals and financial organizations. These attacks often request the potential victim to update their personal information, usernames, passwords, or bank details.

Phishing attacks have become prevalent in the digital age as attackers constantly evolve their methods to deceive unsuspecting users. For example, early phishing attacks involved poorly written emails that were easily identifiable. However, attackers began personalizing emails to target specific people or organizations. They also cloned legitimate emails to trick their potential victims. These attacks evolved to reconnaissance, malware deployment, and data exfiltration. As they continue to evolve, traditional detection techniques, such as simple filters and rule-based systems, cannot effectively tackle modern phishing threats.

The Anti-Phishing Working Group (APWG) observed a record high of 1,270,883 total phishing attacks in the third quarter (Q3) of 2022, an increase from previous years [5, 7]. As of December 2023, about 44.2 million USD was stolen by cybercriminals through phishing attacks [8]. These trends highlight the limitations of current anti-phishing methods and strategies to successfully detect and mitigate phishing attacks. Various methods have been devised to detect these attacks as the hazard increases at a rapid pace. Hence, AI can identify patterns and anomalies indicative of phishing attacks. Furthermore, AI can analyse URLs, email content, and user behavior patterns to identify phishing attempts. The increasing sophistication of phishing tactics necessitates the development of advanced detection mechanisms to proactively identify and mitigate these threats.

Consequently, artificial intelligence (AI) has emerged as a promising avenue for enhancing the efficacy of phishing detection systems. AI utilizes machine learning (ML) and deep learning (DL) algorithms for phishing detection by recognizing patterns and anomalies within datasets [5]. Machine learning uses algorithms and data to predict [9]. This approach employs algorithms to learn and create data-driven predictions or choices [5]. The model can be trained based on existing datasets to learn the properties and patterns of phishing attacks and improve its accuracy in detecting threats [2]. Meanwhile, deep learning uses artificial neural networks to mimic the learning process like that of a human brain [10]. Unlike traditional rule-based systems that rely on predefined heuristics, AI-based methods can learn from data, adapt to new threats, and improve over time [5, 11]. This is significant where attacks constantly evolve their attack methods and threats to avoid detection. AI-driven methods have a more significant advantage over traditional methods based on accuracy and precision issues [12]. The traditional methods require meticulous manual calculations or sequential programming and are slower and less accurate [13]. Contrarily, AI-driven methods execute the processing of large volumes of data in real time and with greater accuracy. AI techniques achieve a much higher level of prediction and prescription capability sophistication than traditional methods. Traditional approaches mainly rely on descriptive analytics, offering fewer insights from historical data [14]. Predictive abilities in traditional systems generally only consist of very simple statistical forecasts. In contrast, techniques driven by AI-driven techniques use more advanced algorithms to predict future trends and recommend actionable strategies [15, 16].

Support vector machines (SVMs) algorithms are widely used for classification, regression, and binary classification problems [17]. They are usually robust in handling high-dimensional data, especially in detecting phishing attacks [18]. Decision trees and random forests build a series of decision rules based on the data's features [19]. They are also effective in detecting phishing attacks as they can be easily interpreted and handle categorical numerical data [20]. Convolutional neural networks (CNN) can effectively analyze visual content and layout and identify subtle differences between legitimate and phishing websites [21]. Hence, AI in phishing detection aims to enhance real-time threat detection and mitigation by improving the accuracy of identifying malicious activities and reducing dependency on human intervention.

However, the features and classification algorithms applied significantly affect the effectiveness of phishing detection. Recently, researchers have shown an increased interest in using AI to detect phishing attacks [9, 18, 22, 23]. Research on the subject has been inconsistent based on the classification techniques, objectives, and dataset used. This study identifies the characteristics and features of AI algorithms applied to detect phishing attacks. The review aims to provide a coherent understanding of the current state of AI-enabled phishing detection and identify areas for future research.

2 Related Work

In the literature, several studies are on detecting, preventing, and mitigating phishing attacks. The use of artificial intelligence in phishing detection is a fast-expanding topic, with numerous studies exploring different techniques and models to improve detection performance.

Machine learning (ML) has been extensively applied to phishing detection with significant success. Jain and Gupta [24] applied machine learning and 19 features using 2141 datasets from PhishTank and Openfish to differentiate between legitimate and fraudulent websites. The algorithm achieved an accuracy of 99.4%. Similarly, Singh et al. [25] employed CNN on a dataset of 36,400 containing legitimate and phishing URLs. They reported an accuracy of 98%. Khemani et al. [14] found that a novel model based on graph neural networks had excellent accuracy rates. The model was applied to four text-size datasets. However, ML may struggle to keep up with the evolving phishing techniques if it is not constantly updated.

Neural networks have been helpful in phishing detection. The study of Ariyadasa et al. [26] achieved an accuracy of 99.1% when using recurrent neural network (RNN) and LSTM algorithms to detect phishing attacks from 10,000 web pages. In the same vein, Rangapur et al. [27] applied RNN to 100,000 legitimate and phishing web pages to achieve 95.79% accuracy. Despite the higher accuracy rates, the neural network requires extensive label data, including phishing and legitimate websites, to train effectively. Also, the datasets need to be regularly updated.

Several authors have explored classifier algorithms for phishing detection. A classifier machine learning algorithm categorizes input data into predefined classes or categories [5]. Joshi and Patternshetti [28] examined different techniques for accurately detecting phishing attacks. Their findings showed that classifiers with feature election provided the best performance. They also found it to be robust. A model for phishing detection using features and classification algorithms was suggested by Ozcan et al. [29]. Kapan et al. [5] applied several features, and classifiers were used to detect phishing attacks. Based on accuracy, precision, and recall performance metrics, they found that the decision tree had the highest F-1 score of 99.0% and an accuracy of 96%.

Machine learning methods were used by Kaur et al. [16] to identify phishing assaults. A deep learning-based system for phishing web page identification was presented by Tang and Mahmoud [30]. Additionally, legitimate and phishing websites have been classified using the multilayer perceptron classifier [31]. However, most studies in this context have relied on the commonly used classifiers [9, 18, 32, 33].

In the literature, few reviews on phishing attack detection have been conducted, offering greater insights into existing methodologies. Zuraiq and Alkasassbe [34] reviewed 18 studies from 2013 to 2018 to identify phishing websites. However, the authors did not consider machine learning approaches to phishing detection. Kunju et al. [35] survey studies from 2007 to 2019 to identify phishing attacks. The authors concluded that many of the proposed solutions were insufficient and need to be revised. Basit et al. [36] conducted a survey consisting of 21 studies on AI-based phishing detection. They suggested that machine learning offers better results than other approaches. Korkmaz et al. [37] limited their literature review on detecting URL-based phishing to five studies. While Catal et al. [38] reviewed 43 studies to assess approaches to phishing detection, their study was limited to only deep learning. The systematic review of Benavides et al. [39], which examined publications on phishing detection studies, was limited to deep learning-based approaches. Thus, these studies had limited focus and did not consider multiple approaches and classifiers in detecting phishing attacks.

3 Method

This study's systematic follows the methodology of similar studies [2, 40, 41]. The methodology includes formulating research questions, identifying relevant databases, data collection, and analysis. The findings are compared and discussed. This systematic literature aims to determine the best approach, dataset, and machine learning algorithm for detecting phishing attacks. This review used the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [42].

3.1 PRISMA Framework

The PRISMA framework provides a transparent, reproducible, and thorough review process. Identification ensures a complete and unbiased search of relevant literature. Screening filters out irrelevant studies quickly to focus on the most pertinent articles. Eligibility thoroughly evaluates each study to ensure it meets inclusion criteria. Finally, inclusion ensures that only the most relevant and high-quality studies related to phishing detection are included in the analysis. It ensures more robust and reliable conclusions. The selected databases and the specified time frame identify the most recent and relevant studies. It aims to understand the current state of AI-based phishing detection research thoroughly. Furthermore, inclusion and exclusion criteria were meticulously selected to maintain the quality and relevance of the review. To conduct a systematic literature review, we followed these steps:

1. Identification: IEEE Xplore, SpringerLink, MDPI, and ScienceDirect were searched using keywords: "phishing attack detection," "AI methods," "machine learning," and "deep learning."
2. Screening: Initial screening was based on titles and abstracts to remove irrelevant studies.
3. Eligibility: Full texts of selected articles were assessed against inclusion and exclusion criteria.
4. Inclusion: Relevant studies meeting all criteria were included in the final review.

Research Questions

1. What are the most common phishing attacks detected?
2. Which algorithms have been used by authors in the studies?
3. Which algorithm has the best accuracy in detecting phishing attacks?
4. What are the most reliable datasets used in the studies?

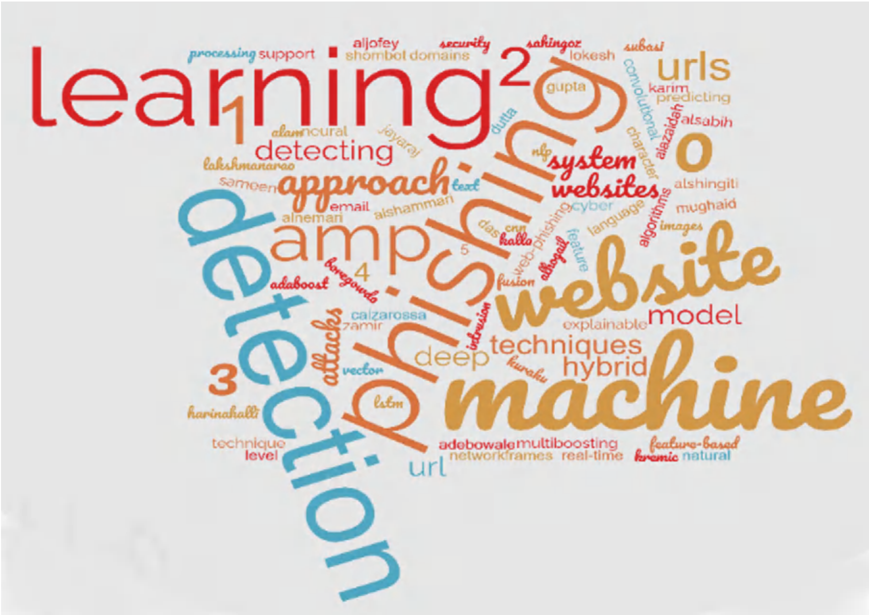


Fig. 1 Word cloud for the keywords of the selected research items

3.2 Important Keywords

The study explored all primary and additional sources of information for the given list of keywords. The search was conducted for papers published from 2019 to 2024. The following keywords were used to search the research items: "phishing attack detection," "AI methods," "classifiers," "features," "machine learning," and "deep learning". The keywords are illustrated in Fig. 1.

3.3 Inclusion and Exclusion Criteria

The inclusion criteria included articles published within the last five years to reflect this context's latest trends, developments, and methodologies. Additionally, it ensures that the review reflects the effectiveness of current and relevant practices rather than outdated methods. Table 1 presents the inclusion and exclusion criteria.

Inclusion Criteria:

- Studies published between 2019–2024
- Peer-reviewed journal articles
- Focus on phishing attack detection using AI methods
- Available in full text.

Table 1 Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
Publication date	2019–2024	Before 2019
Peer-reviewed	Yes	No
Focus	Phishing attack detection using AI methods	Non-AI methods for phishing detection
Full text	Available	Not available
Language	English	Non-english
Publication type	Journal articles	Preprints, reviews, book chapters

Exclusion Criteria:

- Studies not in English
- Studies published before 2019
- Preprints, reviews, and book chapters
- Studies focusing on non-AI methods for phishing detection.

3.4 Search of Relevant Documents

A comprehensive and systematic review is required. Several factors were considered when selecting the databases to ensure the review's comprehensiveness, relevance, and quality. The most relevant electronic databases were explored. These databases were selected based on the reputation focus, our area, and the diversity of research articles. quality, and relevance. They also provide recent studies in AI-based phishing attack detection. Four databases, which include reputable peer-reviewed journals, were selected, and the search yielded the following results:

(a) IEEE Xplore: 166 articles
(b) SpringerLink: 193 articles
(c) MDPI: 52 articles
(d) ScienceDirect: 121 articles.

During the initial search, 532 articles were identified. The duplicates were deleted from search results loaded into Mendeley Reference Manager Software. After removing duplicates and screening for relevance, 45 articles were deemed eligible for full-text review. Only studies that matched the inclusion criteria were included in the review. The full-text articles that satisfied the criteria were retained. The review included papers on phishing attraction detection using AI published between 2019 and 2024 in English. The final inclusion comprised 21 articles (Fig. 2).

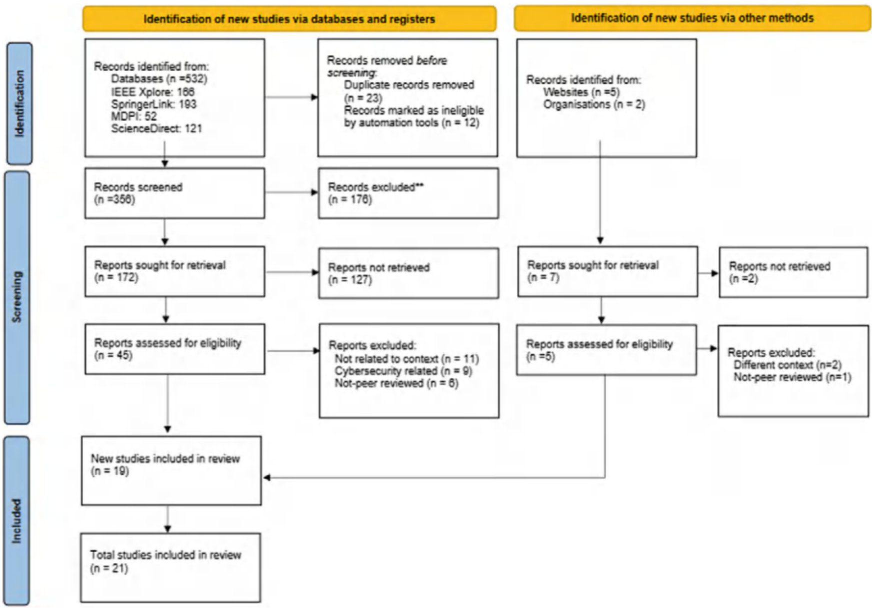


Fig. 2 Study screening and selection

3.5 Quality Evaluation

According to Safi and Singh [2], three categories were used for quality parameters by applying inclusion and exclusion criteria, as shown in Table 2. It aims to focus on the requirements set for a literature review. Hence, each stage (1, 2, and 3) had to be satisfactory before proceeding to the next stage.

Table 2 Quality evaluation criteria

Quality evaluation stage 1
Systematic review is related to phishing attack detection? Yes/No The subject of the study is phishing attack detection techniques Included in the systematic literature review? Yes/No The papers are research articles, experimental or case studies? Yes/No
Quality evaluation stage 2
Study objective is to identify various phishing attack detection techniques? Yes/No Articles highlighted various types of data utilised in detecting phishing attacks? Yes/No SLR revealed different algorithms implemented in phishing attack detection? Yes/No Studies reported the highest accuracy archived by different algorithms? Yes/No
Quality evaluation stage 3
Were the findings of the studies stated clearly? Yes/No Are data from the studies suitable for comparative analysis? Yes/No

4 Results

4.1 Attributes of the Included Research

Most papers, 90.5% (n = 19/21), were published in academic journals and 9.5% (n = 2/21) in conference papers. Figure 3 illustrates the distribution of the included articles per year. Interestingly, the data in Fig. 3 shows that despite a decline in published papers from 2020 to 2022, there has been a significant increase in the number of publications in recent years in this context. This suggests a greater interest, indicating that the area is attracting the attention of scholars. From the data in Fig. 3, most papers were published between 2022 and 2024 (n = 14).

A closer look at the studies shows that most employed Random Forest (RF) compared to other classifiers. Likewise, most studies that applied RF had the highest accuracy [9, 43–47]. Although CNN had a higher accuracy of 99.2%, it is only applied in two studies [10, 48]. Other studies proposed hybrid models using several classifiers [22, 23, 32, 33, 48–50]. However, not all hybrid models had a higher accuracy. The classifiers with the highest accuracy, including the hybrid models, are shown in Table 3. Nevertheless, the hybrid classifiers were not tested across multiple studies. Hence, they may not be applicable in all real-life scenarios. Meanwhile, Random Forest showed a consistently high prediction accuracy rate across several studies.

Recall measures the model's effectiveness in detecting positive instances in the dataset [51]. Table 4 analyses the classifiers with the highest recall scores in the studies. The Linear Regression model has the greatest recall rate, 99.8%. followed by CNN at 99.2%, DTs at 98.8%, XG Boost at 98.8%, and ANFIS at 98.3%.

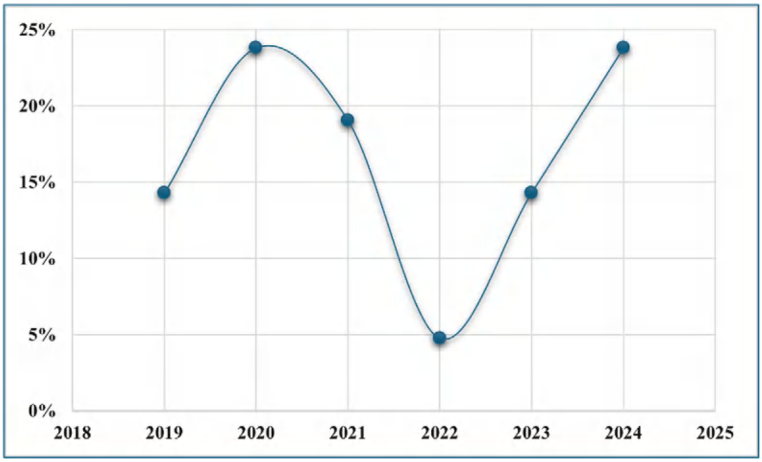


Fig. 3 Publications according to year

Table 3 Classifiers with highest accuracy in the selected studies

Classifier	Average accuracy (%)	Study
XG Boost	99.89	[49]
PILFER	99.50	[32]
Hybrid LSD	98.12	[50]
CNN	98.10	[48]
Random Forest	97.22	[9, 43–47]

Table 4 Classifiers with highest recall (true positive) in the selected studies

Classifier	Average accuracy (%)	Study
Linear Regression	99.8	[32]
CNN	99.2	[10]
XG Boost	98.8	[52]
ANFIS	98.3	[53]
Hybrid	98.3	[23]

Precision measures the frequency of correct predictions for the positive class [51]. The classifiers with the highest precision scores in the studies are depicted in Table 5. The Linear Regression model had the greatest recall rate at 99.7%, while CNN came in second at 99.0%, RF at 99.0%, Naïve Bayes at 98.8%, and a hybrid model at 98.5%.

F1 represents the performance metric for classification. It utilizes recall and precision to determine the model's accuracy [51]. Table 6 shows the classifiers with the highest F-1 scores in the studies analysed. The Linear Regression model also has the highest F1 score at 99.8%, followed by CNN at 99.2%. The F-1 scores for the Hybrid model, ANFIS, and RF classifiers are 98.6%, 98.3%, and 98.0%, respectively.

Table 7 shows a consistent trend across the studies using URL-based, content-based, and behavioral features to detect phishing websites. In the selected studies, the features adopted range from 15 to 40, indicating a comprehensive range of attributes considered essential for effective phishing detection. Standard features include accuracy, recall, precision, classification time, and F1-score. URL-based features include analyzing the URL structure, length, presence of suspicious characters, and domain age. These features are essential as phishing URLs often differ from the usual

Table 5 Classifiers with highest precision score in the selected studies

Classifier	Average accuracy	Study
Linear Regression	99.7%	[32]
CNN	99.0%	[10]
Random Forest (RF)	99.0%	[47]
Naïve Bayes	98.8%	[54]
Hybrid Model	98.5%	[23]

Table 6 Classifiers with F-1 score in the selected studies

Classifier	Average accuracy (%)	Study
Linear Regression	99.8	[32]
CNN	99.2	[10]
Hybrid Model	98.6	[23]
ANFIS	98.3	[53]
Random Forest (RF)	98.0	[47]

patterns of legitimate URLs. Content-based features include specific keywords, the webpage structure, and metadata. Such features help identify the content specific to phishing sites that mimic legitimate websites. However, there are usually slight differences between them. To enhance detection accuracy, behavioral features include user interaction patterns, such as click-through rates and navigation behaviors.

4.2 Strengths and Weaknesses of AI Techniques in the Studies

Regarding the strengths and weaknesses of the various AI techniques, the hybrid model utilizes several phishing indicators to enhance precision and scalability (Das Gupta et al. 2024). Its ability to handle large datasets makes it suitable for real-world applications. It is also effective in scenarios where data volumes are substantial. However, the hybrid model's complexity is a significant implementation and maintenance challenge. The high computational requirements can slow down the process, affecting its effectiveness in real-time applications. Explainable machine learning for phishing feature detection is user-friendly and offers insights into model decisions, increasing users' trust and transparency [43]. Nevertheless, emphasis on explainability can result in trade-offs with detection accuracy and reduced performance. SVMs have excellent generalization capabilities, which enables them to perform well on unnoticed data [18]. Additionally, SVMs are adaptable as they can be applied to various data types like text and images. However, training SVMs with large datasets can be computationally intensive. It can result in longer training times and higher resource consumption.

In the case of utilising multiple machine learning techniques for phishing detection, this approach allows the models to adapt to evolving phishing tactics [9, 50, 55, 56]. This method also improves detection rates over time, reduces false positives, and increases security [22, 44, 46]. Despite this, the approach requires high computational and data resources. It can be a significant barrier to implementation. It is also complex and requires constant updates to remain effective. Furthermore, machine learning models require continuous retraining with new data to remain effective. NLP techniques' versatility allows their application to various text data types, such as URLs

Table 7 Selected studies in the review

Studies	Applied approach	Used algorithm	Datasets	Main findings	Limitations
Das Guptta et al. [49]	Hybrid feature-based phishing detection	Random Forest (RF) Decision Tree (DT) SVM Logistic Regression XG Boost	3,000 legitimate URLs and 3000 phishing URLs	XG Boost has accuracy of 99.89% The proposed method is very effective	Limited dataset size may affect generalizability Third party-dependent features will increase the complexity of the proposed method
Calzarossa et al. [43]	Explainable machine learning for phishing detection	Explainable ML techniques	5,000 phishing websites and 5000 legitimate website URLs	Model based on RF model and feature selection has 96% accuracy	Requires balance between accuracy and interpretability
Shombot et al. [18]	Predictive model for phishing attacks	Support vector machine (SVM)	1,350 data from the Phishtank data archive	Polynomial the function performed better with 84.5% accuracy	Older datasets are employed, which can reduce the performance of the detector with real-time URLs Performance may vary with different datasets
Alazaidah et al. [55]	Website phishing detection	Machine learning techniques	1,248 datasets	FilteredClassifier –F had 90.76% accuracy InfoGainAttributeEval feature had the highest accuracy of 96.58%	Needs larger datasets for robust performance
Jayaraj et al. [56]	Intrusion detection based on phishing using CDG-g	Machine learning techniques	3,000 records	Effective integration of phishing detection into intrusion systems; 96.80% accuracy	Accuracy can vary with different attack vectors and based on the data features

(continued)

Table 7 (continued)

Studies	Applied approach	Used algorithm	Datasets	Main findings	Limitations
Alnemari and Alshammari [9]	Domain phishing detection	ANN, SVM DT RF algorithms	11,055 UCI phishing domains dataset	RF had an accuracy of 96.86% and 97.3% before and after applying the normalisation technique	Examine other machine learning algorithm techniques for phishing domains
Kalla and Kuraku [57]	URL phishing detection	LR RF DT KNN SVM Linear SVC Naïve Bayes	400,000 legitimate URLs and 150,000 phishing URLs	LR, Linear SVC, and Naïve Bayes had an accuracy of 90%	Requires extensive preprocessing of URL data Did not use ensemble models, which can improve the reliability and accuracy of detection
Alshingiti et al. [10]	Deep learning-based phishing detection	CNN LSTM LSTM-CNN	20,000	CNN algorithm has the highest attribute of 99.2%	CNN is relevant to URL In datasets with similar characteristics The website's URL status is not detected by the model
Karim et al. [50]	Hybrid machine learning phishing detection	DT LR RF Navïe Bayes Gradient Boosting KNN SVC Hybrid LSD model	11,054 website datasets from Kaggle	Hybrid LSD model with Canopy feature selection had an accuracy of 98.12%	It may require complex feature extraction

(continued)

Table 7 (continued)

Studies	Applied approach	Used algorithm	Datasets	Main findings	Limitations
Mughaid et al. [32]	Deep learning phishing detection	SVM Boosted DT LR Neural Network DT Averaged perceptron	6,950 non-phishing emails and 860 phishing emails	PILFER had an accuracy of 99.5%	Feature selection techniques need further improvement to handle new methods by cybercriminals
Gururaj Harinahalli Lokesh and Goutham BoreGowda [44]	Phishing website detection	One Class SVM Linear SVC KNN DT RF	MillerSmiles and Phish Tank archives	Random Forest had the highest accuracy of 96.87%	Dataset size and diversity issues
Alhogai and Alsabih [23]	Email phishing detection using a model based on deep learning GCN algorithms	ML and NLP Techniques	3,685 phishing emails and 4894 legitimate emails	The model has an accuracy of 98.2% Recall at 98.3%, precision at 98.5%, and F-1 score at 98.6%	Performance dependent on email characteristics
Lakshmanarao et al. [45]	Phishing website detection	LR Naive Bayes DRT RF Adaboost Gradient Boosting	11,055 UCI dataset	Model combined RF and DT classifier with Gradient Boosting to attain 97% accuracy and recall of 97.0%	Fusion approach complexity may affect implementation Model tested on a single dataset

(continued)

Table 7 (continued)

Studies	Applied approach	Used algorithm	Datasets	Main findings	Limitations
Dutta [22]	Phishing website detection	LURL	7,900 malicious and 5800 legitimate sites from Phishtank dataset	The proposed model showed an accuracy of 97.4% and F-1 score of 96.4%	It may require large datasets for accurate detection The model was not applied to a more extensive network
Alam et al. [46]	Phishing attacks detection	Principal Component Analysis (PCA) RF DT	Dataset from the Kaggle website	RF had an accuracy of 97% and precision of 96.9% DT had recall of 93.8% and F-1 at 90.8% RF had less variance and unable to handle the over-fitting problem	Data quality and preprocessing can impact results It did not employ a convolution neural network (CNN) to test the intrusion detection system The dataset is not specified
Subasi and Kremic [58]	Phishing website detection	Adaboost MultiBoosting SVM KNN ANN RF CART Rotation Forest REP Tree RT C4.5	Publicly available dataset	Adaboost + SVM achieved 97.61% accuracy and F-1 score of 97.6%	Need to tailor the ML algorithms to detect phishing attacks in mobile devices quickly

(continued)

Table 7 (continued)

Studies	Applied approach	Used algorithm	Datasets	Main findings	Limitations
Aljofey et al. [48]	Phishing detection from URL using character-level CNN	MNB LR GNB RF XGB DNN CNN	318,642 empirical URLs	The model achieved an accuracy of 95.02% CNN had recall at 98.6%, precision at 98.6%, and F-1 at 98.6%	CNN is more suited for detecting images rather than text
Sameen et al. [33]	Real-time phishing URL detection	AI Techniques	100,000 phishing and regular URLs	PhishHaven achieved 98.00% accuracy SVM had precision of 97.63% and F-1 score of 97.64%	Real-time processing demands high performance
Sahingoz et al. [47]	URL-based phishing detection	Naive Bayes RFKNN Adaboost K-star SMO DT	36,400 legitimate URLs and 37,175 phishing URLs from Phishtank	Random Forest with NLP-based features has a 98.98% accuracy, precision at 99.0%	Needs additional time for training the system The model needs additional time to check the features and may not be suitable for real-time detection

(continued)

Table 7 (continued)

Studies	Applied approach	Used algorithm	Datasets	Main findings	Limitations
Adebowale et al. [53]	Web-phishing detection by classifying text, image, and frame features	ANFIS KNN SVM Quadratic	13,000 datasets from 2 universities	ANFIS showed an accuracy of 98.3%, recall at 98.3%, precision at 98.3%, and F-1 score at 98.3% Good integrated solution	Integration of multiple features can increase complexity Test deep learning algorithms to detect phishing websites
Zamir et al. [54]	Phishing website detection based on diverse ML algorithms	RF NN Naive Bayes SVM Bagging KNN	11,055 URLs from Kaggle	Stacking1 (RF + NN + Bagging) outperformed other classifiers with 97.4% accuracy, 98.1% recall, and 97% F-1 score	The outcome is dependent on the quality and reliability of the extracted features

and emails [57]. This approach can improve detection accuracy. However, NLP techniques can be computationally intensive and require substantial processing power. Additionally, large, labelled datasets are required to train the models effectively. Integrating machine learning and NLP effectively analysed and detected phishing content within emails [23]. The explainability of NLP models makes results more understandable to users. Nonetheless, the process is computationally intensive and requires significant processing power and memory.

Advanced deep learning techniques like CNN, LSTM, and LSTM-CNN result in high detection accuracy, especially with URLs [59]. Deep learning can automatically adapt to new phishing tactics, improving its long-term efficacy [32]. However, like ML techniques, deep learning models are complex and require high computational power. Moreover, these complex models are challenging to design, train, and interpret. Phishing website detection using a novel machine learning fusion approach showed higher accuracy and robustness [45]. This was because several algorithms were combined. Regardless, the complexity of integrating various machine learning techniques requires regular maintenance and additional resources. Also, it can be demanding when implemented in smaller organisations. Meanwhile, CNNs allow for detailed analysis of URL structures to detect phishing [48]. The approach accounted for subtle differences and patterns that traditional methods might overlook, resulting in higher accuracy.

Similarly, CNNs can be computationally demanding and complex, making them challenging to interpret and maintain. URL-based detection can identify phishing attempts that exploit specific URL patterns [47]. Its efficiency and high accuracy make it suitable for practical implementation in various environments [33]. However, real-time systems require robust infrastructure and significant resources to operate effectively. In addition, reliance on URL-based features may not detect phishing attempts that do not follow typical URL patterns.

4.3 Trends of the Research Concerning Approaches Used

4.3.1 Hybrid Models

Models combining more than one feature set (URL, HTML content, textual data) have consistently shown better performance than single-feature models [49, 50]. Das Guptta et al. [49] combined URL and HTML content features and applied multiple algorithms, including Random Forest, Decision Trees, Support Vector Machines, Logistic Regression, and XG Boost. The XG Boost algorithm showed the best accuracy, at 99.89%, although this was again done on a small dataset, possibly not representative, and involved third-party features that can increase complexity. This trend emphasizes the critical value of multiple, varied feature incorporations for enhanced detection accuracy. Figure 4 displays the accuracy of each algorithm.

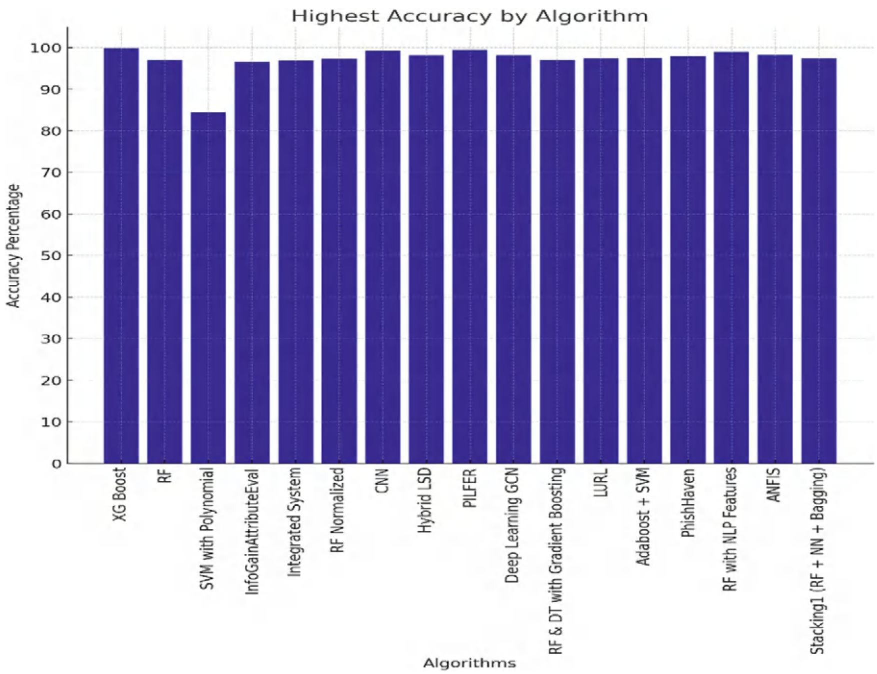


Fig. 4 Accuracy of algorithms

4.3.2 Deep Learning

Deep learning has gained significant attention because it can easily extract complex features from data. Also, it has demonstrated a superior performance in various classification tasks. Two standard deep learning models in phishing detection include long short-term memory (LSTM) and convolutional neural networks (CNNs), and networks are two of the most used deep learning models in phishing detection. Similarly, studies in CNNs and LSTMs for phishing detection showed better results than traditional machine-learning methods [10, 32]. CNN and LSTM networks were applied to 20,000 URLs to obtain 99.2% accuracy. The studies showed one of the highest accuracies than others.

4.3.3 Dataset Diversity

The diversity and representativeness of training datasets are critical to an effective detection model. SVMs using data from the Phishtank archive offered the best performance on the polynomial function with an 84.5% accuracy rate. However, the PhishTank archive was used in four studies; it is an older dataset that may compromise performance with real-time URLs. This is because phishing threats and malicious

Table 8 Datasets used in the selected studies

Dataset	Frequency	Study
Malicious and legitimate sites	5	[22, 33, 43, 49, 57]
PhishTank data archive	4	[18, 22, 44, 47]
Publicly available dataset	4	[48, 55, 56, 58]
Kaggle	3	[46, 50, 54]
Phishing and non-phishing emails	2	[23, 32]
Phishing domains dataset	1	[9]
MillerSmiles	1	[44]
UCI dataset	1	[45]
Datasets from 2 universities	1	[53]

URLs constantly evolve and require constant updates. Five studies employed datasets containing malicious and legitimate sites for phishing threat detection using classifiers. Other datasets used in the studies include Kaggle, MillerSmiles, and UCI. One study utilized datasets from two universities in the United States, while two studies accessed phishing and non-phishing email datasets. Table 8 presents the various datasets employed in the selected studies.

4.3.4 Natural Language Processing

Kuraku and Kalla [57] used NLP and machine learning. While they reported high accuracy in phishing URL detection, the system had some specific values missing, making the reproduction challenging. The approach was reportedly time-consuming due to the extensive preprocessing of URL data.

4.3.5 Classifiers

The analyses considered various approaches that had been used concerning the classifiers. Most studies used only one classifier [9, 18, 22, 43]. Four studies employed two classifiers [23, 45, 57, 58]. Likewise, four studies employed three classifiers [10, 44, 47, 56]. Two studies included five classifiers [48, 49]. Furthermore, one study applied six [50], seven [33], and eight [54] classifiers. Additionally, one study utilized ten classifiers [32]. The distribution of these papers reflects the diversity and complexity of approaches to phishing detection using machine learning techniques. Figure 5 depicts the number of classifiers used in each study.

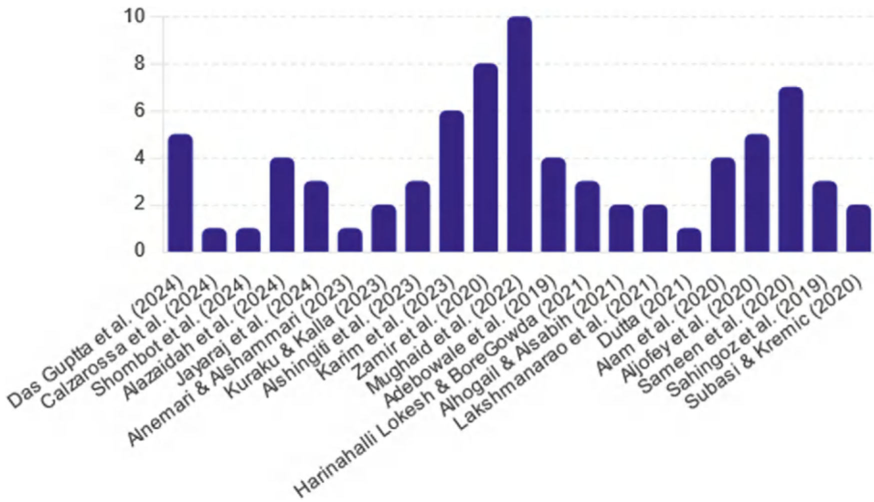


Fig. 5 Classifiers for each study

5 Discussion

This systematic review focused on the most accurate prediction algorithms for detecting phishing attacks. Table 3 shows the prediction accuracy of the classification methods in predicting phishing attack performance from 2019 to 2024. The findings from this review reveal that Random Forest has been the most used method in predicting phishing attacks as it provides reliable and consistent results. Furthermore, XG Boot, CNN, LURL, MK, PILFER, and Hybrid LSD produced good results. Although CNN had an average accuracy of 98.1%, it is more suited to detecting images than text. The linear regression classifier had the highest precision, recall, and F-1 scores at 99.7%, 99.8%, and 99.8%, respectively. Meanwhile, CNN had the second highest scores with precision at 99.0%, recall at 99.2%, and F-1 at 99.8%.

Hybrid approaches that combine multiple AI techniques were explored by several studies to improve phishing detection performance. These studies leverage the strengths of different models to achieve better accuracy and robustness. For example, since ML can handle feature extraction while DL is excellent at classification, combining ML and DL models can provide complementary benefits [60]. The findings suggest that combining multiple approaches in a hybrid model can improve detection accuracy and offer a more comprehensive analysis of phishing attempts. Hence, authors have shifted to applying hybrid approaches to detect phishing attacks by combining different classifiers.

The literature review showed that some studies used ensemble learning techniques by combining multiple ML models to improve detection performance. The results were promising. This supports previous studies, which showed that aggregating several ML modes, such as boosting and bagging methods, can enhance the accuracy of phishing detection systems and make them more robust [61, 62]. This study concludes that this technique helps mitigate the weaknesses of individual models for more reliable phishing detection.

In this literature review, studies using more features tend to report better accuracy in phishing detection. Their overall performance was determined using recall, precision, and F1 scores. The literature stated that including more diverse features can enhance a model's ability to generalize across different types of phishing attacks [63]. Hence, it can identify various phishing attacks, is more resource-efficient, and improves detection rates. Correspondingly, Mohammed et al. [64] found that combining URL and content-based features significantly improved detection rates than a single feature. Studies highlighted the importance of hybrid features, including behavioral futures, as they can efficiently detect sophisticated phishing attempts that elude traditional URL and content checks [65]. Although Alnemari and Alshammari [9] study used fifteen features, less than other studies in the systematic review, they achieved good accuracy. However, their approach may lack the robustness of higher-feature models. The literature [15, 66] suggested that while more features usually improve performance, computational efficiency can balance feature count for greater accuracy.

Regarding classifiers, most studies with single classifiers had limited results. Other studies used multiple classifiers, including Random Forest, SVM, KNN, and Naive Bayes, to complement the strengths of each classifier and improve detection accuracy. Research by Taha [67] demonstrated that ensemble methods, which combine multiple classifiers, typically outperform single classifiers in phishing detection tasks. The findings also align with other studies [68, 69] Sheneamer et al. [70] emphasized the effectiveness of hybrid models combining deep learning techniques like CNNs with traditional classifiers to improve detection rates. Thus, employing multiple classifiers for phishing detection can lead to more accurate and reliable phishing detection [54]. Additionally, they can reduce overfitting and provide a more complete analysis of potential phishing threats. However, they are more complex and require additional resources for computation. Although properly tuning the multiple classifiers' parameters can prevent overfitting, this can be challenging. Furthermore, there may be disagreements among classifiers. Accordingly, researchers must consider these issues and the characteristics of each classifier when selecting them for investigations.

5.1 Research Questions

1. **What are the most common types of phishing attacks detected?**

It was identified as part of a systematic literature review that the most common phishing attacks are fraudulent URLs, spoofed websites, and infected URLs in emails using malware links. The studies mainly targeted URLs because the evolving phishing attacks utilize them.

2. **Which algorithms have been used by authors in the studies?**

The study revealed that the researchers regularly use 40 different algorithms to detect phishing attacks in literature. The algorithms employed in the selected studies comprise traditional machine learning, including RF, DT, SVM, Logistic Regression, and Naive Bayes. Deep Learning consists of CNN, LSTM, LSTM-CNN. The hybrid Models include XG Boost, PhishHavan, Hybrid LSD, PILFER, ANFIS, and Adaboost+SVM.

3. **What algorithm gives us the best accuracy in detecting phishing attacks?**

The highest level of classification accuracy among the reviewed research studies was 99.89% using the XG Boost. The classifier also had a recall score of 98.8%. CNN also had high accuracy at 99.2%, recall at 99.2%, precision at 99.0%, and F-1 at 99.2%. Previous studies confirm that CNN has better prediction accuracy results than traditional machine learning algorithms (Safi and Singh [2]). Similarly, Wei et al. [71] found CNN to have the highest accuracy in detecting phishing attacks, among others. Despite RF having a lower accuracy of 97.22% than XG Boost and CNN, it was utilized in more studies and with consistently high accuracy across the studies. The RF classifier had a precision score of 99.0% and an F-1 score of 98.0. This confirms the findings of Gupta et al. [52] that RF is the best-performing traditional machine-learning algorithm.

4. **What are the most reliable datasets used in the studies?**

Phishtank and Kaggle datasets were more reliable and were used in several studies. Studies show that these two datasets are widely used in academic research and practical studies on phishing attack detection using AI [31, 36]. However, some studies customized their datasets to train a fully robust model. Many studies also compiled datasets from different sources to ensure representation and diversity.

Overall, the findings from these recent studies align well with past research, confirming that combining diverse features and multiple classifiers enhances the robustness and accuracy of phishing detection systems [54, 70]. However, researchers should consider the benefits and limitations of using multiple classifiers. The characteristics of each classifier in a hybrid model must be carefully considered to avoid classifier disagreement or overfitting. Additionally, while more features can improve performance, feature count needs to be balanced with computational efficiency for greater accuracy.

6 Conclusion

AI methods have shown significant promise in enhancing phishing detection accuracy. This study reviewed the latest machine learning methods used to predict phishing attacks. The analysis identified twenty-one relevant studies in this context. The reviewed identified a growing trend in the number of studies in this field. Based on the accuracy metric, this study concluded that XG Boost is an excellent predictor of phishing threats, followed by CNN. However, Random Forest was the most consistent predictor of phishing attacks across several studies. The findings suggest that the most suitable classifier should be selected, and that the importance of various variables be assessed in order to enhance phishing detection.

The review showed a need for a real-time phishing detection system to mitigate phishing threats quickly. This study found that hybrid models like PhishHaven were successful. With the increasing prevalence of phishing attacks, organizations must develop strong and efficient methods for detecting phishing attempts. Our research indicates that companies need to enhance their approaches for effectively detecting and responding to phishing attacks utilizing appropriate classifiers and feature sets with machine learning to protect themselves and their customers.

Despite the advancements in AI-based phishing detection, several challenges remain. One challenge is the evolving nature of phishing attacks. Cybercriminals continually develop new techniques to evade detection, requiring constant updates and improvements to detection models [16]. Additionally, the imbalance in phishing datasets, where legitimate instances far outnumber phishing instances, poses a challenge for training accurate models. Synthetic data generation and data augmentation techniques can help address this issue [72]. Future research should optimize learning parameters and explore hybrid approaches to enhance detection accuracy and scalability. Future research may examine more potential variables and classifiers. Finally, some datasets utilized in the studies were unbalanced, hence disparities in the number of legitimate and phishing sites. Therefore, research should develop more robust and adaptive phishing detection models to handle imbalanced datasets. Likewise, integrating multiple data sources, such as user behavior, social media activity, and network traffic data, can provide a more comprehensive view of phishing attacks and improve detection accuracy.

References

1. Alkhalil, Z., Hewage, C., Nawaf, L., Khan, I.: Phishing Attacks: A Recent Comprehensive Study and a New Anatomy, (2021). https://doi.org/10.3389/fcomp.2021.563060.
2. Safi, A., Singh, S.: A systematic literature review on phishing website detection techniques. Journal of King Saud University—Computer and Information Sciences. 35, 590–611 (2023). https://doi.org/10.1016/j.jksuci.2023.01.004.
3. Paliath, S., Qbeitah, M.A., Aldwairi, M.: Phishout: Effective phishing detection using selected features. In: Proceedings of the 2020 27th International Conference on Telecommunications, ICT 2020. Institute of Electrical and Electronics Engineers Inc. (2020). https://doi.org/10.1109/ICT49546.2020.9239589.
4. Nakamura, A., Dobashi, F.: Proactive phishing sites detection. In: Proceedings - 2019 IEEE/WIC/ACM International Conference on Web Intelligence, WI 2019. pp. 443–448. Association for Computing Machinery, Inc (2019). https://doi.org/10.1145/3350546.3352565.
5. Kapan, S., Sora Gunal, E.: Improved Phishing Attack Detection with Machine Learning: A Comprehensive Evaluation of Classifiers and Features. Applied Sciences (Switzerland). 13, (2023). https://doi.org/10.3390/app132413269.

6. Stojnic, T., Vatsalan, D., Arachchilage, N.A.G.: Phishing email strategies: Understanding cybercriminals' strategies of crafting phishing emails. SECURITY AND PRIVACY. 4, (2021). https://doi.org/10.1002/spy2.165.
7. Thomopoulos, G.A., Lyras, D.P., Fidas, C.A.: A systematic review and research challenges on phishing cyberattacks from an electroencephalography and gaze-based perspective. Pers Ubiquitous Comput. (2024). https://doi.org/10.1007/s00779-024-01794-9.
8. Loh, P.K.K., Lee, A.Z.Y., Balachandran, V.: Towards a Hybrid Security Framework for Phishing Awareness Education and Defense. Future Internet. 16, (2024). https://doi.org/10.3390/fi16030086.
9. Alnemari, S., Alshammari, M.: Detecting Phishing Domains Using Machine Learning. Applied Sciences (Switzerland). 13, (2023). https://doi.org/10.3390/app13084649.
10. Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q.E.U., Saleem, K., Faheem, M.H.: A Deep Learning-Based Phishing Detection System Using CNN, LSTM, and LSTM-CNN. Electronics (Switzerland). 12, (2023). https://doi.org/10.3390/electronics12010232.
11. Abdillah, R., Shukur, Z., Mohd, M., Murah, T.M.Z.: Phishing Classification Techniques: A Systematic Literature Review, (2022). https://doi.org/10.1109/ACCESS.2022.3166474.
12. Stadlmann, C., Zehetner, A.: Comparing AI-Based and Traditional Prospect Generating Methods. Journal of Promotion Management. 28, 160–174 (2022). https://doi.org/10.1080/10496491.2021.1987973.
13. Moro-Visconti, R., Cruz Rambaud, S., López Pascual, J.: Artificial intelligence-driven scalability and its impact on the sustainability and valuation of traditional firms. Humanit Soc Sci Commun. 10, (2023). https://doi.org/10.1057/s41599-023-02214-8.
14. Khemani, B., Patil, S., Kotecha, K., Tanwar, S.: A review of graph neural networks: concepts, architectures, techniques, challenges, datasets, applications, and future directions. J Big Data. 11, (2024). https://doi.org/10.1186/s40537-023-00876-4.
15. Chormunge, S., Jena, S.: Efficient feature subset selection algorithm for high dimensional data. International Journal of Electrical and Computer Engineering. 6, 1880–1888 (2016). https://doi.org/10.11591/ijece.v6i4.9800.
16. Kaur, R., Gabrijelčič, D., Klobučar, T.: Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion. 97, (2023). https://doi.org/10.1016/j.inffus.2023.101804.
17. Roy, A., Chakraborty, S.: Support vector machine in structural reliability analysis: A review, (2023). https://doi.org/10.1016/j.ress.2023.109126.
18. Shombot, E.S., Dusserre, G., Bestak, R., Ahmed, N.B.: An application for predicting phishing attacks: A case of implementing a support vector machine learning model. Cyber Security and Applications. 2, (2024). https://doi.org/10.1016/j.csa.2024.100036.
19. Kayode-Ajala, O.: Applying Machine Learning Algorithms for Detecting Phishing Websites: Applications of SVM, KNN, Decision Trees, and Random Forests. International Journal of Information and Cybersecurity. 6, (2022).
20. Aburrous, M., Hossain, M.A., Dahal, K., Thabtah, F.: Intelligent phishing detection system for e-banking using fuzzy data mining. Expert Syst Appl. 37, 7913–7921 (2010). https://doi.org/10.1016/j.eswa.2010.04.044.
21. Ji, F., Lee, K., Koo, H., You, W., Choo, E., Kim, H., Kim, D.: Evaluating the Effectiveness and Robustness of Visual Similarity-based Phishing Detection Models. (2024).
22. Dutta, A.K.: Detecting Phishing Websites Using Machine Learning Technique. PLoS One. 16, (2021). https://doi.org/10.1371/journal.pone.0258361.
23. Alhogail, A., Alsabih, A.: Applying machine learning and natural language processing to detect phishing email. Comput Secur. 110, (2021). https://doi.org/10.1016/j.cose.2021.102414.
24. Jain, A.K., Gupta, B.B.: Towards detection of phishing websites on client-side using machine learning based approach. Telecommun Syst. 68, 687–700 (2018). https://doi.org/10.1007/s11235-017-0414-0.
25. Singh, S., Singh, M.P., Pandey, R.: Phishing detection from URLs using deep learning approach. In: Proceedings of the 2020 International Conference on Computing, Communication and Security, ICCCS 2020. Institute of Electrical and Electronics Engineers Inc. (2020). https://doi.org/10.1109/ICCCS49678.2020.9277459.

26. Ariyadasa, S., Fernando, S., Fernando, S.: Detecting phishing attacks using a combined model of LSTM and CNN. International Journal of Advanced and Applied Sciences. 7, 56–67 (2020). https://doi.org/10.21833/ijaas.2020.07.007.
27. Rangapur, A., Kanakam, T. P. D.: Phish-Defence: Phishing Detection Using Deep Recurrent Neural Networks. (2021).
28. Aniruddha N. Joshi, Tanuja R. Patternshetti: Phishing Attack Detection using Feature Selection Techniques. In: International Conference on Communication and Information Processing (2019).
29. Ozcan, A., Catal, C., Donmez, E., Senturk, B.: A hybrid DNN–LSTM model for detecting phishing URLs. Neural Comput Appl. 35, 4957–4973 (2023). https://doi.org/10.1007/s00521-021-06401-z.
30. Tang, L., Mahmoud, Q.H.: A Deep Learning-Based Framework for Phishing Website Detection. IEEE Access. 10, 1509–1521 (2022). https://doi.org/10.1109/ACCESS.2021.3137636.
31. Aljammal, A.H., taamneh, S., Qawasmeh, A., Salameh, H.B.: Machine Learning Based Phishing Attacks Detection Using Multiple Datasets. International Journal of Interactive Mobile Technologies. 17, 71–83 (2023). https://doi.org/10.3991/ijim.v17i05.37575.
32. Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., Elsoud, E.A.: An intelligent cyber security phishing detection system using deep learning techniques. Cluster Comput. 25, 3819–3828 (2022). https://doi.org/10.1007/s10586-022-03604-4.
33. Sameen, M., Han, K., Hwang, S.O.: PhishHaven - An Efficient Real-Time AI Phishing URLs Detection System. IEEE Access. 8, 83425–83443 (2020). https://doi.org/10.1109/ACCESS.2020.2991403.
34. AlMaha Abu, Z., Mouhammd, A.: Review: Phishing Detection Approaches. In: 2nd International Conference on New Trends in Computing Sciences (ICTCS) :Proceedings (2019).
35. Merlin .V. Kunju, Esther Dainel, Heron Celestie Anthony, Sonali Bhelwa: Evaluation of Phishing Techniques Based on Machine Learning. IEEE (2019).
36. Basit, A., Zafar, M., Liu, X., Javed, A.R., Jalil, Z., Kifayat, K.: A comprehensive survey of AI-enabled phishing attacks detection techniques, (2021). https://doi.org/10.1007/s11235-020-00733-2.
37. Mehmet Korkmaz, Ozgur Koray Sahingoz, Banu Diri: Feature Selections for the Classification of Webpages to Detect Phishing Attacks: A Survey. In: HORA 2020 : 2nd International Congress on Human-Computer Interaction, Optimization and Robotic Applications Proceedings (2020).
38. Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., Shukla, S.: Applications of deep learning for phishing detection: a systematic literature review. Knowl Inf Syst. 64, 1457–1500 (2022). https://doi.org/10.1007/s10115-022-01672-x.
39. Benavides, E., Fuertes, W., Sanchez, S., Sanchez, M.: Classification of Phishing Attack Solutions by Employing Deep Learning Techniques: A Systematic Literature Review. In: Smart Innovation, Systems and Technologies. pp. 51–64. Springer Science and Business Media Deutschland GmbH (2020). https://doi.org/10.1007/978-981-13-9155-2_5.
40. Kitchenham, B., Pretorius, R., Budgen, D., Brereton, O.P., Turner, M., Niazi, M., Linkman, S.: Systematic literature reviews in software engineering-A tertiary study, (2010). https://doi.org/10.1016/j.infsof.2010.03.006.
41. Singh, S., Beniwal, H.: A survey on near-human conversational agents, (2022). https://doi.org/10.1016/j.jksuci.2021.10.013.
42. Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., Group, P.: Preferred reporting items for systematic reviews and meta-analyses : the PRISMA statement. PLoS Med. 6, 1–8 (2009). https://doi.org/10.1136/bmj.b2535.
43. Calzarossa, M.C., Giudici, P., Zieni, R.: Explainable machine learning for phishing feature detection. Qual Reliab Eng Int. 40, 362–373 (2024). https://doi.org/10.1002/qre.3411.
44. Gururaj Harinahalli Lokesh, Goutham BoreGowda: Phishing website detection based on effective machine learning approach. Journal of Cyber Security Technology. 5, 1–14 (2021). https://doi.org/10.1080/23742917.2020.1813396.

45. Lakshmanarao, A., Surya Prabhakara Rao, P., Bala Krishna, M.: Phishing website detection using novel machine learning fusion approach. In: Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021. pp. 1164–1169. Institute of Electrical and Electronics Engineers Inc. (2021). https://doi.org/10.1109/ICAIS50930.2021.9395810.
46. Alam, M.N., Sarma, D., Lima, F.F., Saha, I., Ulfath, R.E., Hossain, S.: Phishing attacks detection using machine learning approach. In: Proceedings of the 3rd International Conference on Smart Systems and Inventive Technology, ICSSIT 2020. pp. 1173–1179. Institute of Electrical and Electronics Engineers Inc. (2020). https://doi.org/10.1109/ICSSIT48917.2020.9214225.
47. Sahingoz, O.K., Buber, E., Demir, O., Diri, B.: Machine learning based phishing detection from URLs. Expert Syst Appl. 117, 345–357 (2019). https://doi.org/10.1016/j.eswa.2018.09.029.
48. Aljofey, A., Jiang, Q., Qu, Q., Huang, M., Niyigena, J.P.: An effective phishing detection model based on character level convolutional neural network from URL. Electronics (Switzerland). 9, 1–24 (2020). https://doi.org/10.3390/electronics9091514.
49. Das Guptta, S., Shahriar, K.T., Alqahtani, H., Alsalman, D., Sarker, I.H.: Modeling Hybrid Feature-Based Phishing Websites Detection Using Machine Learning Techniques. Annals of Data Science. 11, 217–242 (2024). https://doi.org/10.1007/s40745-022-00379-8.
50. Karim, A., Shahroz, M., Mustofa, K., Belhaouari, S.B., Ramana, S., Joga, K.: Phishing Detection System through Hybrid Machine Learning Based on URL. IEEE Access. (2023). https://doi.org/10.1109/ACCESS.2017.DOI.
51. Peretz, O., Koren, M., Koren, O.: Naive Bayes classifier – An ensemble procedure for recall and precision enrichment. Eng Appl Artif Intell. 136, (2024). https://doi.org/10.1016/j.engappai.2024.108972.
52. Gupta, V.K., Gupta, A., Kumar, D., Sardana, A.: Prediction of COVID-19 confirmed, death, and cured cases in India using random forest model. Big Data Mining and Analytics. 4, 116–123 (2021). https://doi.org/10.26599/BDMA.2020.9020016.
53. Adebowale, M.A., Lwin, K.T., Sánchez, E., Hossain, M.A.: Intelligent web-phishing detection and protection scheme using integrated features of Images, frames and text, (2019). https://doi.org/10.1016/j.eswa.2018.07.067.
54. Zamir, A., Khan, H.U., Iqbal, T., Yousaf, N., Aslam, F., Anjum, A., Hamdani, M.: Phishing web site detection using diverse machine learning algorithms. Electronic Library. 38, 65–80 (2020). https://doi.org/10.1108/EL-05-2019-0118.
55. Alazaidah, R., Al-Shaikh, A., AL-Mousa, M.R., Khafajah, H., Samara, G., Alzyoud, M., Al-Shanableh, N., Almatarneh, S.: Website Phishing Detection Using Machine Learning Techniques. J Stat Appl Probab. 13, 119–129 (2024). https://doi.org/10.18576/jsap/130108.
56. Jayaraj, R., Pushpalatha, A., Sangeetha, K., Kamaleshwar, T., Udhaya Shree, S., Damodaran, D.: Intrusion detection based on phishing detection with machine learning. Measurement: Sensors. 31, (2024). https://doi.org/10.1016/j.measen.2023.101003.
57. Kalla, D., Kuraku, S.: Phishing Website URL's Detection Using NLP and Machine Learning Techniques. Journal on Artificial Intelligence. 5, 145–162 (2023). https://doi.org/10.32604/jai.2023.043366.
58. Subasi, A., Kremic, E.: Leveraging AI and machine learning for societal challenges, cas 2019 comparison of adaboost with multiboosting for phishing website detection. In: Procedia Computer Science. pp. 272–278. Elsevier B.V. (2020). https://doi.org/10.1016/j.procs.2020.02.251.
59. Alshanqiti, A., Namoun, A.: Predicting student performance and its influential factors using hybrid regression and multi-label classification. IEEE Access. 8, 203827–203844 (2020). https://doi.org/10.1109/ACCESS.2020.3036572.
60. Castiglioni, I., Rundo, L., Codari, M., Di Leo, G., Salvatore, C., Interlenghi, M., Gallivanone, F., Cozzi, A., D'Amico, N.C., Sardanelli, F.: AI applications to medical images: From machine learning to deep learning, (2021). https://doi.org/10.1016/j.ejmp.2021.02.006.
61. Yaman, E., Subasi, A.: Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification. Biomed Res Int. 2019, (2019). https://doi.org/10.1155/2019/9152506.

62. Junior, B., Nicoletti, M. do C.: An iterative boosting-based ensemble for streaming data classification. Information Fusion. 45, 66–78 (2019). https://doi.org/10.1016/j.inffus.2018.01.003.
63. Bhargava R. Maddireddy, Bharat R. Maddireddy: Business & Social Sciences AI-Based Phishing Detection Techniques: A Comparative Analysis of Model Performance. Unique Endeavor in Business & Social Sciences. 01, (2022).
64. Mohammed, S.Y., Aljanabi, M., Mijwil, M.M., Ramadhan, A.J., Abotaleb, M., Alkattan, H., Albadran, Z.: A Two-Stage Hybrid Approach for Phishing Attack Detection Using URL and Content Analysis in IoT. In: BIO Web of Conferences. EDP Sciences (2024). https://doi.org/10.1051/bioconf/20249700059.
65. Omar, A.R., Taie, S., Shaheen, M.E.: From Phishing Behavior Analysis and Feature Selection to Enhance Prediction Rate in Phishing Detection. (2023).
66. Perkash, A., Shaheen, Q., Saleem, R., Rustam, F., Villar, M.G., Alvarado, E.S., Torre Diez, I. de la, Ashraf, I.: Feature optimization and machine learning for predicting students' academic performance in higher education institutions. Educ Inf Technol (Dordr). (2024). https://doi.org/10.1007/s10639-024-12698-9.
67. Taha, A.: Intelligent ensemble learning approach for phishing website detection based on weighted soft voting. Mathematics. 9, (2021). https://doi.org/10.3390/math9212799.
68. Adnan, M., Imam, M.O., Javed, M.F., Murtza, I.: Improving spam email classification accuracy using ensemble techniques: a stacking approach. Int J Inf Secur. 23, 505–517 (2024). https://doi.org/10.1007/s10207-023-00756-1.
69. Feng, W., Gou, J., Fan, Z., Chen, X.: An ensemble machine learning approach for classification tasks using feature generation. Conn Sci. 35, (2023). https://doi.org/10.1080/09540091.2023.2231168.
70. Sheneamer, A.M., Halawi, M.H., Al-Qahtani, M.H.: A hybrid human recognition framework using machine learning and deep neural networks. PLoS One. 19, (2024). https://doi.org/10.1371/journal.pone.0300614.
71. Wei, W., Ke, Q., Nowak, J., Korytkowski, M., Scherer, R., Woźniak, M.: Accurate and fast URL phishing detector: A convolutional neural network approach. Computer Networks. 178, (2020). https://doi.org/10.1016/j.comnet.2020.107275.
72. Maradana Durga Venkata Prasad, Srikanth T: Multi-Entity Real-Time Fraud Detection System using Machine Learning: Improving Fraud Detection Efficiency using FROST-Enhanced Oversampling. Journal of Electric Systems. 20, 1380–1394 (2024).