

CONFERENCE PROCEEDINGS

NATIONAL SMART INNOVATION CONFERENCE 2025

THEME

AI AND THE FUTURE OF COMPUTING:
INNOVATIONS IN RESEARCH, PRACTICE
AND PEDAGOGY

DATE
SEPTEMBER
16 - 18 2025

TIME
9AM - 5PM
DAILY

VENUE
DAY 1 & 2: NOUN, ABUJA
DAY 3: MIVA OPEN UNIVERSITY, ABUJA



Table of Contents

Hybrid Deep Q-Learning and ACO Approach for Dynamic Cloud Resource Optimization	11
Comparative Analysis of Hyperparameter Tuning and Ensemble Methods for Optimizing Machine Learning Predictions in Education	26
Digital Forensics in Mobile Devices on Resources, Techniques, Challenges, And Open Issues: A Systematic Literature Review	43
Prediction of Cryptocurrency (Bitcoin) Price using Machine Learning Algorithms	66
Artificial Intelligence in Human Resource Management: Challenges and Future Directions	88
Identifying Financial Fraud Using Explainable Stacked Ensembles	106
Motion of Earth's Artificial Satellite in the Earth-Moon system around Axial Equilibrium Points under effects of Radiation, Perturbations and Mass Variations in the Robe's R3BP	126
Adoption and Usability of Blended Learning in Secondary Schools in Nigeria: Case Study, Federal Capital Territory	145
Adewole, D. J.	145
Artificial Intelligence – Powered Cyber Warfare: Resources, Methods, Weaponisation, Forensics, Threat Intelligence and Defenses: A Systematic Review	165
Part-of-Speech Tagging for English Discourse Analysis: Foundations and Applications	188
A Systematic Literature Review on Financial Fraud Detection in E-Commerce Systems: Implementation Techniques, Weaknesses, Challenges, and Research Directions	206
A Systematic Literature Review: Machine and Deep Learning Network-Based Intrusion Detection System	225
A Systematic Review on Construction Block Estimation System Using a Machine Learning Tool: Benefits, Challenges, and Future Directions	243
The Use of Technology in Warehouse Management Systems: A Review	265
Balancing Progress and Protection: Regulatory and Ethical Dimensions of AI in Education and ODL	283
Factors Influencing Ethical AI-Generated Content (AIGC) Utilization Among Computer Education Students in a Northern Nigerian University	323
Impact of AI Tools on Academic Writing and Research Practices in Nigerian Institutions	339
Artificial Intelligence Enabled Ransomware: A Systematic Review	360
AI-Driven Portfolio Optimisation Using Reinforcement Learning	384
Personalised Learning Recommender for Open and Distance Learning	408
Regex-Based Divide and Conquer Approach for Parsing ATM Electronic Journal Logs	422
From Scepticism to Synergy: Building Confidence in AI Adoption with Human-Centered Institutional Frameworks	453

A Comprehensive Exploratory Data Analysis of Climatic Drivers of Floods and Droughts Forecasting in Taraba State, Nigeria	472
A Soulbound Non-Fungible Tokens-Based Certificate Verification System for Nigerian Universities	491
Leveraging GeoAI for Mapping and Forecasting Open Distance Learning	511
Educational Needs Among Youth in Nasarawa State, Nigeria	511
Entropy with Decision Tree-Based Anomaly Detection for Mitigating OpenFlow-Based Denial of Service Attacks in Software-Defined Networks	530
A Review of Behavioural Machine Learning Models for Intelligent Detection of Social Engineering Attacks: Issues, Trends, and Research Directions	551
Comparative Analysis of Query Retrieval and Performance of Xml, Sqlite, and Flat File Database	580
Artificial Intelligence for Post-Quantum Cryptography and Post-Quantum Cryptography for Artificial Intelligence: Dual Pathways to Quantum-Resilient Digital Security	598
Comparative Analysis of Virtual Reality and Traditional Pedagogy: A Mini-Systematic Literature Review	627

Preface

Artificial Intelligence (AI) is no longer a distant promise but a living force shaping education, economies, governance, healthcare, and everyday life. The future of computing through AI is not about imitation but about innovation rooted in the realities of Nigeria and Sub-Saharan Africa.

These proceedings of the 2025 National Smart Innovation Conference are jointly organized by the Faculty of Computing, National Open University of Nigeria (NOUN), and the School of Computing, Miva Open University, Nigeria. The theme of the conference is: *“Artificial Intelligence and the Future of Computing: Innovations in Research, Practice and Pedagogy”*, with a focus on Artificial Intelligence in Education and Open and Distance Learning; Software Engineering and Intelligent Systems; FinTech and Computational Economics; Cybersecurity and Trustworthy Computing; Communication, IoT and Emerging Technologies; as well as Educational Technology and Pedagogy.

We are very pleased with the response to the call for papers. Submissions received were well aligned with the conference's focus on AI and its profound implications for the future of computing in research, practice, and pedagogy, particularly comparative analysis in Nigeria and across Sub-Saharan Africa.

A total of 70 papers were received, of which 40 were accepted for presentation and publication after a double-blind peer review conducted through a standard electronic submission system by experts and professionals across the globe. Each paper received a minimum of two reviews. We deeply appreciate the efforts of our dedicated reviewers for investing their time and expertise. Without their commitment and professionalism, the success of these proceedings would not have been possible. The

quality of the accepted papers reflects the pride of both the authors and reviewers, who together have contributed to the advancement of knowledge in this field.

The academic staff of the Faculty of Computing, NOUN, and the School of Computing, Miva Open University, are pleased to host this conference as a platform for vibrant ideas, collaboration, and creativity among some of the brightest minds shaping the integration of AI in education and other allied areas of human endeavour.

We extend our sincere appreciation to the Vice-Chancellor of the National Open University of Nigeria, Professor Olufemi Peters, and the Vice-Chancellor of Miva Open University, Professor Olutayo Arulogun; to Digital Encode (our co-sponsor); and to the keynote speakers, guest speakers, session chairs, authors, reviewers, participants, and students for their invaluable contributions to the success of this conference.

We wish all participants a fruitful conference and a pleasant stay in Abuja.

Editors

Prof. O. M. Olaniyi

Prof. G. Obunadike

Dr. E. E. Ogbuju

Dr. M.O. Lawrence

Reviewers:

Dr. Emmanuel Oluwanbo Asani, Landmark University, Omu-aran

Dr. Martha Ozohu Musa, University of Port Harcourt, Port Harcourt.

Dr Roseline Oluwaseun Ogundokun, Landmark University, Omu Aran

Oluwatoyin Adelakun-Adeyemo, Bingham University, Karu, Nasarawa

Dr. Halleluyah Oluwatobi Aworinde, Bowen University, Iwo.

Dr. Isah Charles Saidu, Baze University, Abuja

Prof. Aderonke A Oni, Covenant University, OTA.

Chioma Isiekwene, Miva Open University, Abuja

Dr. Noah Akande, Nile University of Nigeria, Abuja

Dr. Emmanuel Igbekele, Landmark University

Dr. Timothy Shamaki, Federal University of Technology, Babura

Prof. Wilson Nwankwo, Delta State University of Science & Technology

Kwale-Ughelli Road, Ozoro, Delta State, Nigeria

Dr Adeola Kolawole, Nigeria Defense Academy, Kaduna

Dr Nuhu Bello, Federal University of Technology, Minna

Dr Alfa Abraham, Confluence University of Science and Technology, Osara

Prof Gerogina Obunadike, NOUN Abuja

Dr Farida Garki, NOUN Abuja

Dr Christana Ezeanya, NOUN Abuja

Dr Okunade Oluwasogo, NOUN Abuja

Dr Wale Adebayo, NOUN Abuja

Dr Egenti Grace, NOUN Abuja

Dr Idirs Sinan, NOUN Abuja



National Smart Innovation Conference

Introduction

It is my pleasure to introduce the proceedings of this conference. The National Smart Innovation Conference 2025 is the first edition of this gathering, born out of a collaboration between the NOUN and Miva Open University. I wish to specially thank my Vice-Chancellor, Prof. Tayo Arulogun, whose visionary leadership inspired me to initiate this conference. I am grateful to the pioneer Dean of the Faculty of Computing at NOUN, Dr. Greg Onwodi, for graciously accepting the idea at its inception. My deepest appreciation also goes to the current Dean, Prof. Olayemi Mikail OLANIYI, for steering and executing the conference with utmost dedication, wisdom, and passion. Let me also extend my heartfelt thanks to the Chairman of the Central Planning Committee, Mr. Jimin Wuese, and the Chairperson of the Local Organizing Committee, Prof. Georgina Obunadike, for their excellent coordination and tireless efforts in ensuring the success of this conference. Their commitment, attention to detail, and collaborative spirit have been instrumental in bringing every aspect of this event to fruition.

We named it the *National Smart Innovation Conference* because of its focus on Artificial Intelligence for national transformation, with this year's edition highlighting the application of AI in computing education within the open and distance learning context.

Day One opens with a grand ceremony featuring welcome addresses from the Vice-Chancellors of NOUN and Miva Open University, followed by keynote speeches from our esteemed speakers. The day peaks with a panel session on AI-driven innovation in education, practical workshops on AI research and prompt engineering, and concludes with interactive poster presentations.

Day Two begins with keynotes on pedagogy in the AI era, followed by three exhibitions showcasing real-world innovations. A panel discussion on AI as a catalyst for computing then leads into multiple paper presentation tracks covering themes such as artificial intelligence, education and ODL, cybersecurity, and curriculum innovation.

We have carefully curated this programme to balance keynote insights, scholarly research, practical workshops, and networking opportunities. Our aim is to provide a platform that redefines the role of AI in education and innovation for Nigeria and the wider world.

These proceedings present good research contributions that extend beyond education into other areas of AI, including natural language processing. While not exhaustive, they reflect the growing depth of scholarship in this field. We are hopeful that this conference will become an annual event, improving in scope and impact with each passing year.

We shall continue to work smarter. Enjoy your proceedings!

Dr. Emeka Ogbuju

Dean, School of Computing, Miva Open University, Abuja

Co-host, National Smart Innovation Conference

HYBRID DEEP Q-LEARNING AND ACO APPROACH FOR DYNAMIC CLOUD RESOURCE OPTIMIZATION

Hybrid Deep Q-Learning and ACO Approach for Dynamic Cloud Resource Optimization

Abdulwasiu A. A.¹, Obunadike G. N.,² and Olanrewaju O. M.³

Computer Science Department, Federal University Dutsinma, Katsina, Nigeria^{1,2,3}

Corresponding author: ibnwasar@gmail.com

Abstract

Cloud computing offers on-demand access to shared resources, but effectively allocating these resources is still a challenge due to fluctuating workloads, high computational expenses, and the slow convergence of current algorithms. Existing methods often struggle to optimize makespan, resource utilization, and energy consumption all at once. To tackle these issues, this study introduces ACO+DQL, a hybrid model that integrates Deep Q-Learning (DQL) for adaptive decision-making with Ant Colony Optimization (ACO) for global search enhancement. The approach was tested with 120 cloudlets assigned to 10, 20, and 50 virtual machines (VMs). The results indicate that ACO+DQL surpasses standalone ACO and DQL, achieving resource utilization rates of 0.85, 0.6678, and 0.5072, makespan times of 97.76 seconds, 68.99 seconds, and 37.80 seconds, and energy consumption figures of 100,789 J, 136,788 J, and 250,234 J, respectively. These results illustrate a scalable, energy-efficient, and adaptable solution for dynamic cloud environments.

Keywords: Cloud Computing, Resource Allocation, Deep Q-Learning, and Ant Colony Optimization

1. Introduction

Cloud computing has become a game-changing approach that allows users to access a shared set of computing resources, like servers, storage, applications, and services, over the internet whenever they need them. Its ability to easily scale and its pay-as-you-go pricing model have made it appealing to both businesses and individuals. In a cloud data center, how resources are allocated is important to ensuring that applications work efficiently and perform well. Typically, new tasks (called cloudlets) are assigned to virtual machines (VMs) in the data center, and how well this assignment is done affects the system's overall performance, energy use, resource efficiency and security (Olanrewaju et al., 2023). Thus, finding the best way to allocate resources is essential for enhancing service quality (QoS) and improving the energy efficiency of cloud systems, meaning that optimization is important (Obunadike et al., 2023).

To tackle this challenge of resource allocation, several metaheuristic and Machine Learning algorithms have been utilized and Metaheuristic is a good approach (Obunadike et al., 2025). Ant Colony Optimization (ACO) was applied for QoS-based scheduling but encountered scalability problems (Sharma et al., 2023). An improved version, Hybrid Weighted Ant Colony Optimization (HWACO), was introduced to enhance convergence but overlooked energy efficiency (Chandrashekar et al., 2023). Recently, reinforcement learning techniques have been investigated, with Xiu et al., (2023) developing MRLCC, which adapted well to changing environments but was primarily focused on CPU resources. Despite these efforts, current methods generally fall short in optimizing energy consumption or experience slow convergence and instability in dynamic settings. This points to the need for a hybrid solution that can

effectively balance exploration and decision-making efficiency as hybrid method prove strong success in optimization.

This research proposes a model for Energy-Efficient Dynamic Resource Allocation that combines Deep Q-Learning and Ant Colony Optimization (ACO+DQL) to fill these gaps. This model enhanced the adaptability of Deep Q-Learning alongside the global search capability of ACO to optimize makespan, energy efficiency, and resource utilization.

2. Related works

Lilhore et al., (2025) developed a Hybrid DRL-IACO-IWWO model aimed at optimizing cloud resource allocation and load balancing. This model combines Deep Reinforcement Learning with Improved Ant Colony Optimization and Water Wave Optimization, utilizing a two-phase strategy where IACO handles initial scheduling and WWO focuses on global optimization, all influenced by feedback from the DRL. Simulation outcomes indicated a scheduling time of 1.25 seconds, an operational cost of \$23.8, energy consumption of 15.6 kWh, and utilization at 92%. The model consistently outperformed ACO-GA, ACO-SMO, and WWO-GA in various conditions, including static, dynamic, and constrained scenarios. However, it suffers from significant computational complexity, substantial training overhead, and limitations in scalability beyond 40 VMs.

Malipatil et al., (2025) presented a framework for energy-efficient workload scheduling in cloud computing that utilizes deep reinforcement learning. Their approach used a Deep Q-Network (DQN) to dynamically assess workload parameters, including execution time, CPU, and memory usage, facilitating adaptive task allocation. Results were impressive, showing diminished latency (15 ms), elevated throughput (500

tasks/sec), 92% load balancing efficiency, 95% resource utilization, and 97% quality of service, while minimizing energy consumption (120 kWh compared to 250 kWh using Round Robin). The model surpassed traditional algorithms such as Round Robin, FCFS, and heuristic approaches. Nonetheless, it is limited by high convergence time and training expenses, which restrict rapid deployment in environments with limited resources.

Mangalampalli et al., (2024) introduced an Adaptive Task Scheduler (ATSIA3C) that employs the Improved Asynchronous Advantage Actor-Critic (IA3C) algorithm for effective task scheduling within multi-cloud environments. This method breaks down incoming tasks into subtasks, groups them based on execution parameters, and assigns them to the appropriate VMs. Simulations using CloudSim with both synthetic and real workloads demonstrated substantial enhancements: a 70.49% reduction in makespan, a 77.42% decrease in resource costs, and a 74.24% decrease in energy consumption compared to existing methodologies. The study illustrates high efficiency when handling complex workloads; however, it faces challenges with high computational complexity and limited adaptability to swiftly changing workloads.

Alhazmi, (2023) introduced a model based on Deep Reinforcement Learning known as Deep Q-Networks with Actor-Critic (DQNAC) for optimal resource allocation and task scheduling in cloud computing. This approach utilized DRL techniques on actual cloud workload data to enhance efficiency. DQNAC demonstrated superior performance over BATS+BAR, GA, HH, and OLOA regarding resource utilization, execution time, response time, and average makespan. Its major contribution is improving the flexibility and efficiency of cloud applications through strategic workload distribution. However, it

relies heavily on high computational resources and may face scalability issues in managing large and diverse cloud environments.

Xiu et al. (2023) proposed MRLCC, a meta reinforcement learning-based method for task scheduling aimed at increasing adaptability in dynamic cloud settings. This technique employs meta reinforcement learning to address challenges related to low sample efficiency and the need for retraining that are prevalent in typical DRL methods. Experimental findings revealed that MRLCC achieved significantly higher utilization rates (up to 0.8479) and shorter makespan compared to five baseline algorithms, such as Round Robin and heuristic methods. The main contribution of this work is its ability to quickly adapt to new environments with minimal gradient updates. However, a limitation is that the model currently only addresses CPU requirements, which restricts its capacity to manage a variety of multi-resource constraints effectively.

Sharma et al. (2022) introduced LBACO, a combination of Genetic Algorithm (GA) and Ant Colony Optimization (ACO), aimed at improving quality of service (QoS) in cloud task scheduling. Their experiments using CloudSim demonstrated that LBACO was effective in reducing execution time, optimizing resource utilization, and lowering costs (ranging from 120 to 260 over 20 iterations) compared to GA and ACO. However, the model is limited by its reliance on static data, a lack of personnel allocation considerations, and difficulties in handling dynamic multi-objective scheduling.

Chandrashekar et al. (2022) developed the Hybrid Weighted Ant Colony Optimization Algorithm (HWACO) to enhance cloud task scheduling by minimizing makespan and costs while boosting efficiency. This methodology improved standard ACO by

adjusting pheromone evaporation rates dynamically based on historical data and utilizing the Construct_ANT and Update_Pheromone functions for optimal solutions. Simulation results indicated that HWACO significantly outperformed QANA, MTF-BPSO, ACO, MIN-MIN, and FCFS, with improvements of up to 57.11% in makespan and 32.94% in cost. Nonetheless, the algorithm entails increased computational overhead due to the weight-based modifications and updates in pheromones.

Zambuk et al., (2021) introduced ACOBF, a hybrid of Ant Colony Optimization and Bacterial Foraging, to optimize cloud task scheduling by minimizing makespan, energy consumption, and waiting times, while enhancing resource utilization. This approach integrated ACO for discovery with bacterial foraging techniques, categorizing tasks into three groups and implementing virtual machine clustering. Testing with 1000 tasks demonstrated that ACOBF outperformed both Genetic Algorithm (GA) and Bacterial Foraging (BF), achieving the best results in terms of makespan and energy usage across 20–200 tasks. A significant drawback is its slow convergence rate and the computational burden due to additional crossover, mutation, and chemotaxis processes.

Rugwiro, Gu, and Ding, (2019) proposed a hybrid model combining Ant Colony Optimization (ACO) and Deep Reinforcement Learning (DRL) for cloud task scheduling, aimed at minimizing completion time and maximizing idle resource utilization. By employing a Binary In-order Traversal Tree for scheduling and CloudSim for simulation, their approach reduced execution time from 250 seconds to 230 seconds for 1000 tasks and improved resource utilization compared to CWSA and ACOPSO. However, the model's effectiveness is highly dependent on parameter tuning for both ACO and DRL, making it sensitive and less reliable under varying workload conditions. Other Machine

learning algorithms have also been used for effective work, such as KNN and SVM (Obunadike et al., 2023).

3. Methodology

This study presented a hybrid approach called ACO+DQL aimed at achieving energy-efficient dynamic resource allocation in cloud computing. It initiates with the preprocessing of 120 diverse cloudlets, which are then allocated across various configurations of Virtual Machines (VMs), including setups with 10, 20, and 50 VMs. This design seeks to strengthen the exploration-exploitation capabilities of Deep Q-Learning (DQL) for adaptive decision-making while improving the global optimization potential of Ant Colony Optimization (ACO) for effective task scheduling and load balancing, as the steps is illustrated in figure 1.

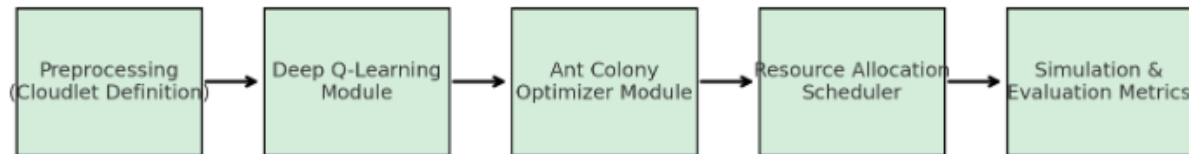


Figure 1: Implementation workflow

From figure 1, in the initial phase, cloudlets were characterized by diverse workloads concerning execution time, length, and resource requirements. These tasks were introduced into the simulation environment, where various configurations and numbers of VMs were deployed. DQL was utilized to model the decision-making process by defining states as task-to-VM mappings and actions as scheduling choices. A carefully crafted reward function aimed to decrease makespan and energy usage while maximizing resource utilization, ensuring that each allocation contributes to overall optimization.

Subsequently, ACO was incorporated during the scheduling phase to enhance the solutions produced by DQL. Each ant symbolizes a possible task-to-VM allocation route, with pheromone updates driven by execution time, energy, and utilization metrics. The collaboration between DQL and ACO facilitates a quicker convergence towards nearly optimal solutions than methods that operate independently. The algorithm was executed using Python, featuring iterative loops that simulate cloud task arrivals, DQL updates through Q-value learning, and ACO pheromone reinforcement.

At the end of the methodology, the system's efficiency was assessed by varying the VMs while maintaining a constant count of 120 cloudlets. The Evaluation performance indicators evaluated included makespan, energy usage, and resource utilization, offering a thorough assessment of the trade-offs in rapidly changing cloud settings. Comparative evaluations with existing baseline algorithms were also conducted to emphasize the efficiency improvements of the introduced ACO+DQL model.

4. Results and Discussion

This section provides a comprehensive analysis of the suggested methodologies across various cloud computing setups. The research examines resource utility, makespan, and energy usage across 120 cloudlets that feature different configurations of virtual machines, specifically those with 10, 20, and 50 VMs, as illustrated in Figures 2 to 4.

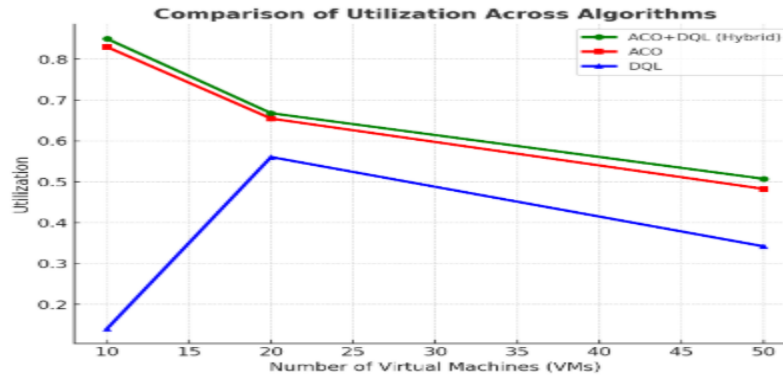


Figure 2: Resource Utilization for 10, 20, and 50 VMs

Figure 2 depicts the resource utilization data from 10, 20, and 50 VMs, which clearly highlight the advantages of the hybrid ACO+DQL algorithm when compared to the individual approaches. With 10 VMs, the hybrid achieved a utilization rate of 0.85, surpassing ACO's 0.83 and DQL's mere 0.14, demonstrating its effectiveness in maximizing resource activity and minimizing idle periods. As the number of VMs increased to 20, the hybrid maintained its lead with a utilization rate of 0.6678, exceeding ACO's 0.6547 and DQL's 0.5605, indicating ongoing success in managing workload distribution. At the 50 VMs level, although there was a general decline in utilization due to complexity, the hybrid still recorded the highest rate of 0.5072 compared to ACO's 0.4825 and DQL's 0.3419. These findings confirm that while ACO excels in heuristic scheduling and DQL is less effective with standalone resource allocation, their integration enhances adaptability, scalability, and sustained resource efficiency in dynamic cloud settings.

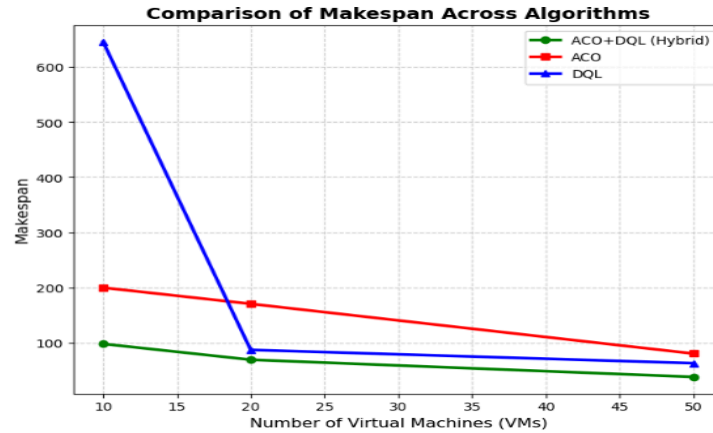


Figure 3: Makespan for 10, 20, and 50 VMs

Figure 3 shows the makespan analysis which indicates that the hybrid ACO+DQL consistently outperformed both the ACO and DQL across all VM configurations. With 10 VMs, the hybrid achieved a makespan of 97.76 seconds, significantly better than ACO's 199.75 seconds and DQL's 645.73 seconds. At 20 VMs, the hybrid recorded 68.99 seconds, outperforming ACO's 170.40 seconds and DQL's 86.87 seconds. For 50 VMs, the hybrid reached a makespan of 37.80 seconds, again lower than ACO's 80.12 seconds and DQL's 62.71 seconds. These findings affirm that the integration of ACO's structured searching capability with DQL's adaptive refinement promotes more rapid and balanced task completion across all configurations.

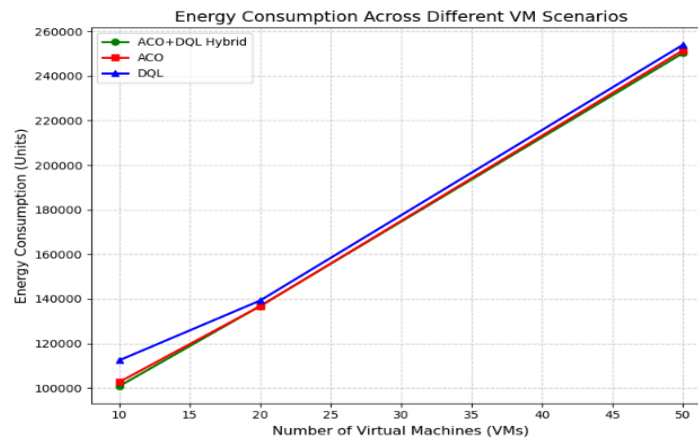


Figure 4: Energy Consumption for 10, 20, and 50 VMs

Figure 4 indicates the energy consumption findings which reveal notable variances among the algorithms relative to VM scales. With 10 VMs, the hybrid ACO+DQL exhibited the least energy usage at 100,789 Joules, while ACO used 102,657 Joules and DQL consumed a higher 112,398 Joules, underscoring the hybrid's efficiency in smaller settings. When increasing to 20 VMs, ACO slightly edged out the hybrid with 136,676 Joules versus 136,788 Joules, with DQL again consuming the most at 139,223 Joules, showing its inefficiency. For the 50 VM scenario, ACO utilized 251,480 Joules closely followed by the hybrid at 250,234 Joules, whereas DQL notably lagged at 253,972 Joules. These results illustrate that while ACO performs well in medium to large-scale environments, the hybrid approach maintains competitive efficiency with minimal overhead, and DQL consistently underperforms due to increased decision-making costs.

5. Conclusion

The suggested ACO+DQL hybrid model effectively optimizes 120 cloudlets using 10, 20, and 50 virtual machines (VMs). It achieves a resource utilization rate of 0.85 compared to standalone algorithms 0.6678, and 0.5072, with makespan times of 97.76 seconds compared to standalone algorithms 68.99 seconds, and 37.80 seconds, along with energy consumption figures of 100,789 joules. These findings demonstrate that integrating Deep Q-Learning with Ant Colony Optimization (ACO) results in task scheduling that is adaptable, scalable, and energy-efficient surpassing the performance of ACO and DQL when used independently.

6. Future work

This research work has contributed using hybrid algorithms which are Deep Q-learning and Ant colony optimizer, however, other algorithms may also be used in future to see their performance. Moreover, the energy consumption needs to be looked into more, though there is little improvement from this research, therefore, means of increasing the performance more should be looked into.

References

- Alhazmi, L. (2023). A deep reinforcement learning-based model for optimal resource allocation and task scheduling in cloud computing. *International Journal on Recent and Innovation Trends in Computing and Communication*, 11(8s), 97–106. <https://doi.org/10.17762/ijritcc.v11i8s.7179>
- Busari, A. I., Obunadike, G. N., & Mukhtar, A. (2025). Power system optimization using hybrid heuristic search algorithm. *BIMA Journal of Science and Technology*, 9(1B), 94–107. <https://journals.gjbeacademia.com/index.php/bimajst/article/view/938>
- Chandrashekar, C., Krishnadoss, P., Kedalu Poornachary, V., Ananthakrishnan, B., & Rangasamy, K. (2023). HWACOA scheduler: Hybrid weighted ant colony optimization algorithm for task scheduling in cloud computing. *Applied Sciences*, 13(6), 3433. <https://doi.org/10.3390/app13063433>
- Lilhore, U. K., Simaiya, S., Rao, K. B. V. B., Maheswara Rao, V. V. R., Sharma, Y. K., Alroobaea, R., Alsufyani, H., Alsafyani, M., & Khan, M. D. M. (2025). Hybrid DRL-enhanced ACO-WWO for efficient resource allocation and load-balancing in cloud computing. *International Journal of Computational Intelligence Systems*, 18(1), 148. <https://doi.org/10.1007/s44196-025-00882-9>

Malipatil, A. R., Paramasivam, M. E., Gulyamova, D., Saravanan, A., Ramesh, J. V. N., Muniyandy, E., Ghodhbani, R., Ali, Y. A., Jabr, B. A., & Abdullaeva, B. S. (2025). Energy-efficient cloud computing through reinforcement learning-based workload scheduling. *International Journal of Advanced Computer Science and Applications*, 16(4), 655–670. <https://doi.org/10.14569/IJACSA.2025.0160483>

Mangalampalli, S., Karri, G. R., Ratnamani, M. V., Mohanty, S. N., Jabr, B. A., Ali, Y. A., Ali, S., & Abdullaeva, B. S. (2024). Efficient deep reinforcement learning-based task scheduler in multi-cloud environment. *Scientific Reports*, 14, 21850. <https://doi.org/10.1038/s41598-024-72774-5>

Obunadike, G., Emeka, O., & Mukhtar, A. (2023). A filter model for text categorization against online hate speeches. *FUDMA Journal of Sciences*, 3(2), 195–200. <https://fjs.fudutsinma.edu.ng/index.php/fjs/article/view/1502>

Obunadike, G., John, A., & Ismaila, I. (2023). Optimization of K-mode algorithm for data mining using particle swarm optimization. *FUDMA Journal of Sciences*, 2(3), 24–33. <https://fjs.fudutsinma.edu.ng/index.php/fjs/article/view/1381>

Olanrewaju, O. M., Abdulwasiiu, A. A., & Abdulhafeez, N. (2023). Enhanced AODV routing protocol to prevent blackhole attack in MANET. *International Journal of Software Engineering and Computer System*, 24(1), 11–18. <https://doi.org/10.15282/ijsecs.9.1.2023.7.0111>

Rugwiro, U., Gu, C., & Ding, W. (2019). Task scheduling and resource allocation based on ant-colony optimization and deep reinforcement learning. *Journal of Internet Technology*, 20(5), 1463–1473. <https://doi.org/10.3966/160792642019092005013>

Sharma, N., Sonal, & Garg, P. (2022). Ant colony-based optimization model for QoS-based task scheduling in cloud computing environment. *Measurement: Sensors*, 24, 100531. <https://doi.org/10.1016/j.measen.2022.100531>

Xiu, X., Li, J., Long, Y., & Wu, W. (2023). MRLCC: An adaptive cloud task scheduling method based on meta reinforcement learning. *Journal of Cloud Computing: Advances, Systems and Applications*, 12, 75. <https://doi.org/10.1186/s13677-023-00440-8>

Zambuk, F. U., Gital, A. Y., Jiya, M., Sabon Gari, N. A., Ja'afaru, B., & Muhammad, A. (2021). Efficient task scheduling in cloud computing using multi-objective hybrid ant colony optimization algorithm for energy efficiency. *International Journal of Advanced Computer Science and Applications*, 12(3), 451–456. <https://doi.org/10.14569/IJACSA.2021.0120354>

COMPARATIVE ANALYSIS OF HYPERPARAMETER TUNING AND ENSEMBLE METHODS FOR OPTIMIZING MACHINE LEARNING PREDICTIONS IN EDUCATION

Comparative Analysis of Hyperparameter Tuning and Ensemble Methods for Optimizing Machine Learning Predictions in Education

Akuma S., Ortese T., Mfena M.K.

Department of Mathematics, Computer Science, and Statistics,

Benue State University, Makurdi – Nigeria

Corresponding author: sakuma@bsum.edu.ng

Abstract

AI tools and applications that are used in different sectors of human endeavour depend on model accuracy to perform their desired functions. Understanding how to fit a model accurately to avoid overfitting is necessary for building future AI tools and applications. Even as this is necessary, the research in optimizing model accuracy is still limited. In this work, we compared the optimization efficiency of hyperparameter tuning and ensemble methods. Data relating to factors affecting students' performance in JAMB examination was sourced from Kaggle.

The data has JAMB score as the target variable and Distance_To_School, Attendance_Rate, Study_Hours_Per_Week, Teacher_Quality, Socioeconomic_Status, Parent_Involvement, School_Type, Parent_Education_Level, Extra_Tutorials, Access_To_Learning_Materials, Parent_Involvement, IT_Knowledge, Assignments_Completed, Age and Gender as the independent variables. The analysis of the results shows a weak correlation between the JAMB score and the variables. To compare the different optimizing models, a baseline model consisting of a Decision Tree and Random Forest was developed. Two other optimized models, consisting of grid search hyperparameter tuning and bagging ensemble methods, were also developed. The evaluation of the models using RMSE and MAE metrics shows that the bagging ensemble optimized model performed better than the baseline models and grid search optimized model, with a lower MAE (40.289%) and RMSE (32.699).

Keywords: Machine learning, optimization, grid search hyperparameter tuning, bagging ensemble, Artificial Intelligence

1. Background

Computing still follows the adage of “garbage in, garbage out”, meaning the quality or type of input data determines the type of output data. Data is an essential factor in Artificial Intelligence since, without it, an AI system will not be able to learn and make future decisions. Data comes in different formats based on the variety, velocity and volume, and is collected over a period of time. In most cases, data used for solving AI problems are not sourced by primary researchers, leading to a possibility of noise in the dataset, like outliers, missing values, wrong data entry, and so on. This noise in data can cause an overfitting or underfitting problem when developing a model (Stephen Akuma & Iqbal, 2018; Tetko et al., 2024). In Educational Data Mining (EDM), information extracted from a dataset with educational relevance can be used for making key decisions in the educational sector. Such information must be accurate to achieve the goals of the stakeholders involved. To develop a high-accuracy model, various techniques have been used by researchers, like feature selection, ensemble, and hyperparameter tuning (Foroozand & Weijs, 2017; Rahmat et al., 2024; Schratz et al., 2019). The ensemble method, like bagging, boosting and stacking, can be used to improve the overall performance of a model by combining predictions from several base models, while hyperparameter tuning determines the ideal hyperparameter settings to enhance the model's performance (A Ilembayo et al., 2024; Gul, 2025).

Recent approaches towards EDM use machine learning classification or regression algorithms and dimensionality reduction to find patterns in educational data (Alharbi et al., 2021; Bellaj et al., 2024; Papadogiannis et al., 2024). In the work by Alharbi et al.

(2021), they optimize model configurations and students' performance hyperparameters using three feature selection algorithms and seven classifiers. Their result show that hyperparameter tuning improves the model for student performance prediction. Hoyos & Caicedo-Castro (2024) research on EDM also found a relationship between predictor variables and students' academic performance. Cuevas-Redondo & Estevez-Bravo (2017) predicted academic performance using Linear Regression and Decision Tree. Their result show that Decision Tree has an error rate of 56% as compared to Linear Regression with an error rate of 64%, signifying that both models may not be good for EDM of small datasets. However, in the research by S. Akuma & Abakpa (2021), Linear Regression was used to predict students' performance from previous academic records. Although the data was relatively small, their result showed an accuracy of 87.84%; a contrast from the findings by Cuevas-Redondo & Estevez-Bravo (2017). Similarly, high accuracy was obtained in research by Kabakchieva (2012) in the same context, with KNN producing 73.59% accuracy and ANN producing 73.59% accuracy. Research on High school students' performance prediction in Brazil was conducted by de Melo et al. (2017). They obtained up to 89% accuracy using Decision Tree algorithm, while Menacho Chiok (2017) obtained an accuracy of 71% for Naïve Bayes algorithm on a similar dataset in the same region. Although the research in EDM is promising and has shown accuracy in prediction with large datasets, this is not always the case with small datasets, especially when there are many features involved. This often leads to an overfitting problem, as shown in Cuevas-Redondo & Estevez-Bravo (2017) above. To improve the accuracy of EDM on a small dataset, feature engineering like hyperparameter tuning and ensemble have been found useful (A Illemobayo et al., 2024). Even though machine learning models are used extensively in EDM, their efficacy is mostly dependent on choosing the best hyperparameters, which is quite challenging and manually done (Kalita et al., 2025).

Sarmah et al. (2024) confirm that when the best hyperparameters are chosen in combination with a predictive model, it produces improved results for most datasets.

Research by Elgeldawi et al. (2021) shows that the grid search method performed better than other hyperparameter tuning methods in improving model accuracy. However, Tetko et al. (2024) cautioned that hyperparameter tuning leads to overfitting when using the same statistical measures. Research by Foroozand & Weijs (2017) on EEF ensemble bagging techniques found effectiveness in the results when 40% of the most informative ensemble members are used. Hasan et al. (2015) infer that the improvement of the accuracy of a model with an ensemble method, such as bagging, varies with different base classifiers, and when a suitable base classifier is selected, the performance of the dataset improves. Opitz & Maclin (1999) supported this assertion with their study on popular ensemble methods. They found that in most situations, bagging is more accurate than a single classifier. In this work, we carried out a comparative examination of two data optimization methods to ascertain which of them is better in improving prediction accuracy for an educational problem. The methods employed include hyperparameter tuning using grid search and an ensemble method using bagging, as shown in Fig. 1 and Fig. 2, respectively.

In machine learning, the grid search method adjusts hyperparameters that determine the best set of parameters for a particular model. For each hyperparameter, a discrete set of potential values must be defined, and then every conceivable combination of these values must be thoroughly evaluated, because a model with poor performance and incorrect results can come from selecting the hyperparameters incorrectly. The grid search process begins by defining the hyperparameter space where a list of specific/discrete values is defined, forming the “grid” of possible hyperparameter combinations. The next phase involves the evaluation of the machine learning model

with each of the combined hyperparameters stated in the grid. Here, standard evaluation metrics like recall and precision with cross-validation are employed. Finally, the best combined hyperparameters that produced the optimal performance are selected for the model (Aktas, 2025).

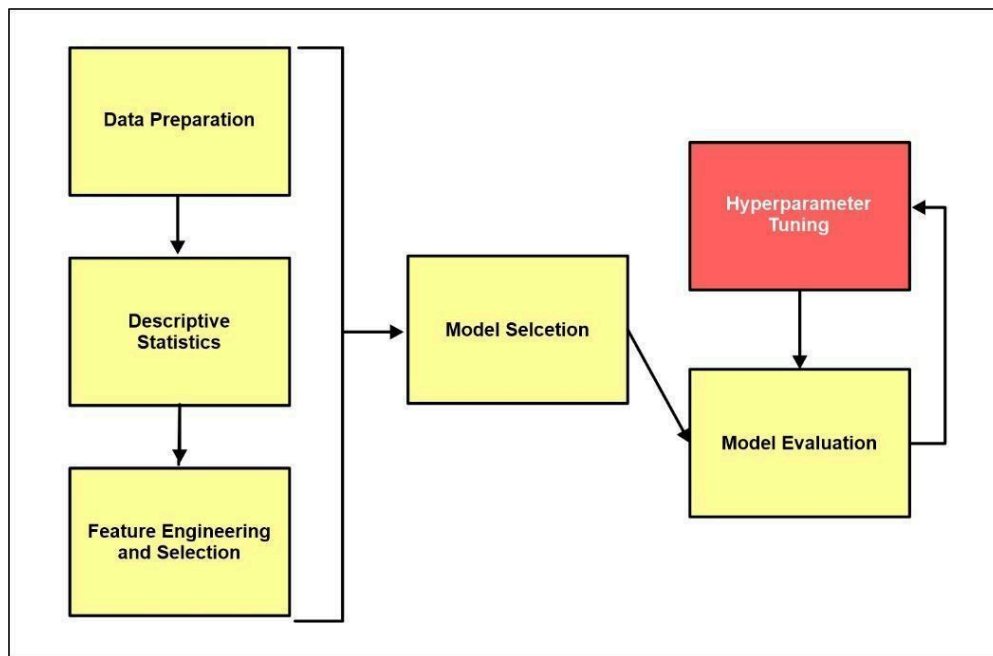


Fig. 1. Hyperparameter tuning process

In the case of Bootstrap Aggregating ensemble learning method, shortened to Bagging, it aims to improve the accuracy of the predicting model by reducing variance. It trains multiple base models independently on different subsets (bags) sampled randomly from the main training data. The process begins by randomly splitting the original training dataset into subsets (bags), then a model like K-Nearest Neighbors, Decision Tree and so on is trained independently on each of the bags. After the base models are trained, a final prediction is derived by combining them (Sarmah et al., 2024).

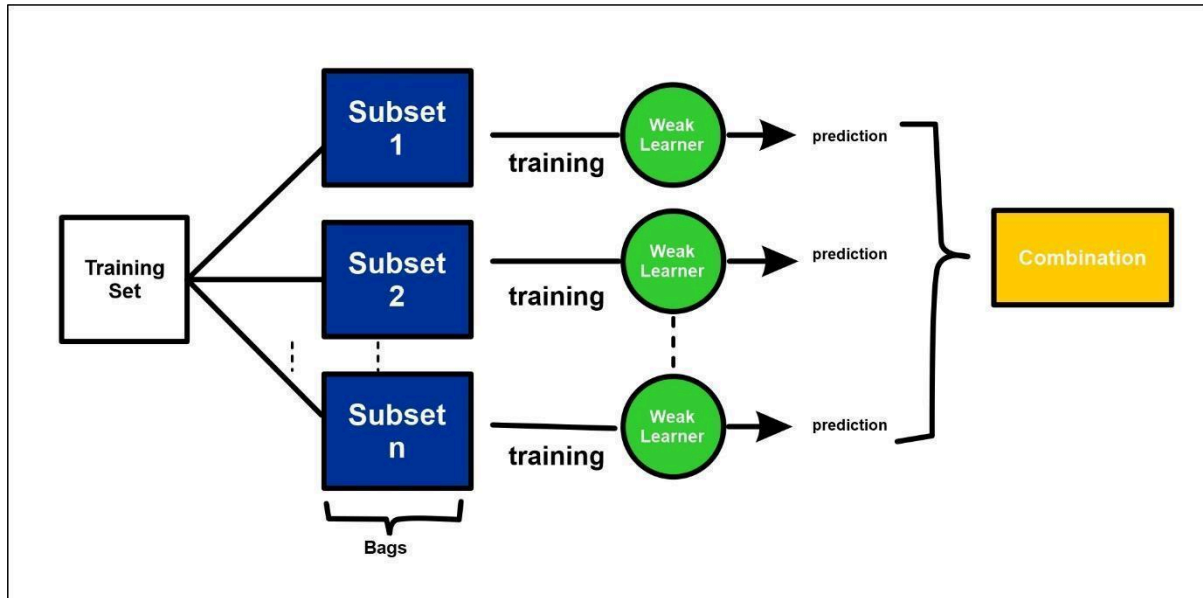


Fig. 2. Ensemble bagging process

2. Method

The methodology used for this research encompassed data preprocessing, model training, performance evaluation, and visualization. The goal of this study is to compare the optimization performance of two techniques, viz, hyperparameter-tuned models and ensemble learning, in improving the predictive model for a real-world educational dataset. Here, we determined the elements influencing student performance. We used an educational dataset from Kaggle that features students' Jamb Scores as well as some behavioural, demographic information (Idowu Adamo, 2024). This dataset was created to forecast student success using data from the 2024 Joint Admissions and Matriculation Board (JAMB) exam. Nigerian universities use the JAMB as a standardised test for admission. We carried out an exploratory data analysis to identify the features available. The dataset comprises 5,000 entries and 17 features, including demographic, educational, and behavioural variables related to student performance in the 2024 JAMB examination. The features used for developing the model are Distance_To_School, Attendance_Rate, Study_Hours_Per_Week,

Teacher_Quality, Socioeconomic_Status, Parent_Involvement, School_Type, Parent_Education_Level, Extra_Tutorials, Access_To_Learning_Materials, Socioeconomic_Status, Parent_Involvement, IT_Knowledge, Assignments_Completed, Age and Gender as independent variables while the JAMB Score is the target variable. We observed that Parent_Education_Level has missing values in 17.82% of the records, so we examined three strategies for handling the missing values, viz, filling with the mode, filling with a random value and removal of rows. Our analysis found that filling the empty cells with related random numbers was the best among the three methods. So, we applied this method to cells with missing data.

To evaluate the efficiency of the grid search hyperparameter tuning approach and bagging ensemble approach, we partitioned the data by splitting it into an 80% training set and a 20% test set using a fixed random seed to ensure reproducibility. To handle heterogeneous feature types, a preprocessing pipeline was constructed. Numerical features were standardized using z-score normalization via Standard Scaler, while categorical features were encoded using one-hot encoding. A Column Transformer was employed to integrate these preprocessing steps within the modelling pipeline.

Three families of models were evaluated, including a baseline Decision Tree Regressor and Random Forest Regressor. Each was embedded within a pipeline that incorporated the preprocessing steps to ensure consistent data handling. A Decision Tree Regressor was subjected to a comprehensive hyperparameter tuning using grid Search. The hyperparameter space included variations in tree depth and node splitting criteria. The parameters tuned were max_depth for controlling the maximum depth of

the tree, and `min_samples_split` for specifying the minimum number of samples needed for splitting the internal node in the tree. For the `max_depth`, we tested for overfitting and underfitting using 3, 5, 10 and 20 values, allowing a wide range of values between shallow trees (3 and 5) to medium trees (10) to deep trees (20). In the case of `min_samples_split`, we tested 2, 5, 10 and 20 values. We found that lower values (2) make the tree very detailed with a potential for overfitting, medium values (5 or 10) are generally more stable, while larger values (20) make simpler trees.

A Decision Tree Regressor with controlled depth as the base estimator was used to instantiate a Bagging Regressor. Bootstrap sampling (bagging) was used to aggregate 50 base learners. For smooth integration, the ensemble model was integrated into the same preprocessing pipeline and evaluated using 10-fold cross-validation.

3. Results and Discussion

This study used a real-world educational dataset to assess how well different regression models predicted JAMB scores. A hyperparameter-tuned Decision Tree employing grid search, a bagging ensemble based on Decision Trees, and the standard Decision Tree with Random Forest regressors as a baseline are among the models studied. Standard assessment metrics, such as the coefficient of determination (R^2), mean absolute error (MAE) in equation 1, and root mean squared error (RMSE) in equation 2, were used to evaluate each model's performance. Furthermore, 10-fold cross-validation was performed to assess each model's stability and generalisability.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

equation 1

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

equation 2

We conducted a correlation analysis to examine the relationship between the independent features and Jamb Score. The result showed a weak correlation between Jamb Score and the independent features, as shown in the heatmap in Fig. 3.

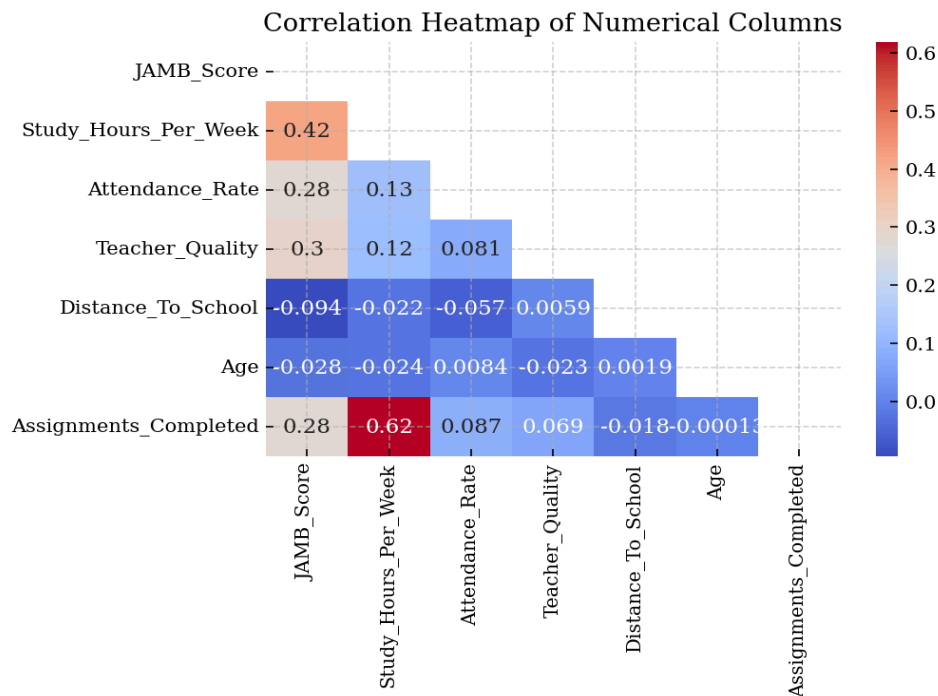


Fig. 3. Correlation heatmap between the independent features and Jamb Score

Additionally, a feature importance analysis was performed using the Random Forest Regressor, wherein feature contributions were extracted and ranked, followed by visualization of the top ten predictors. The results are summarized in Table 1. Among the models tested, Decision Tree (bagging) achieved the best overall performance with an RMSE of 40.289 and R^2 of 0.327 on the test set. It also exhibited the lowest cross-validation RMSE mean (39.611). To enhance interpretability, visual analytics were conducted. Bar plots were generated for comparative model performance across

RMSE, MAE, and R^2 metrics, as shown in Fig. 4. This outcome suggests that bagging can significantly improve model performance on a dataset as compared to grid search hyperparameter tuning.

Table 1. Model Comparison

Model	RMSE	MAE	R^2	CV RSME
Decision Tree	57.580	45.443	0.374	57.547
Random Forest	40.617	33.001	0.316	39.779
Decision Tree (Grid Search)	42.376	34.483	0.256	41.581
Decision Tree (Bagging)	40.289	32.699	0.327	39.611

On the other hand, the baseline Decision Tree Regressor performed the worst, with an RMSE of 57.58 and a negative R^2 of -0.374, indicating poor generalization and potential overfitting. Its cross-validation RMSE also reflected this instability. Random Forest, a commonly robust ensemble method, achieved moderate results (RMSE = 40.61, R^2 = 0.316), slightly worse than Decision Tree (Bagging). This suggests that the ensemble approach prevents overfitting compared to the single tree. Decision Tree (Grid Search) also achieved a significant improvement over the default tree (RMSE = 42.38, R^2 = 0.256). The results, as presented, suggest that modest improvements from ensemble and tuning strategies signify the potential for improving model accuracy on a dataset.

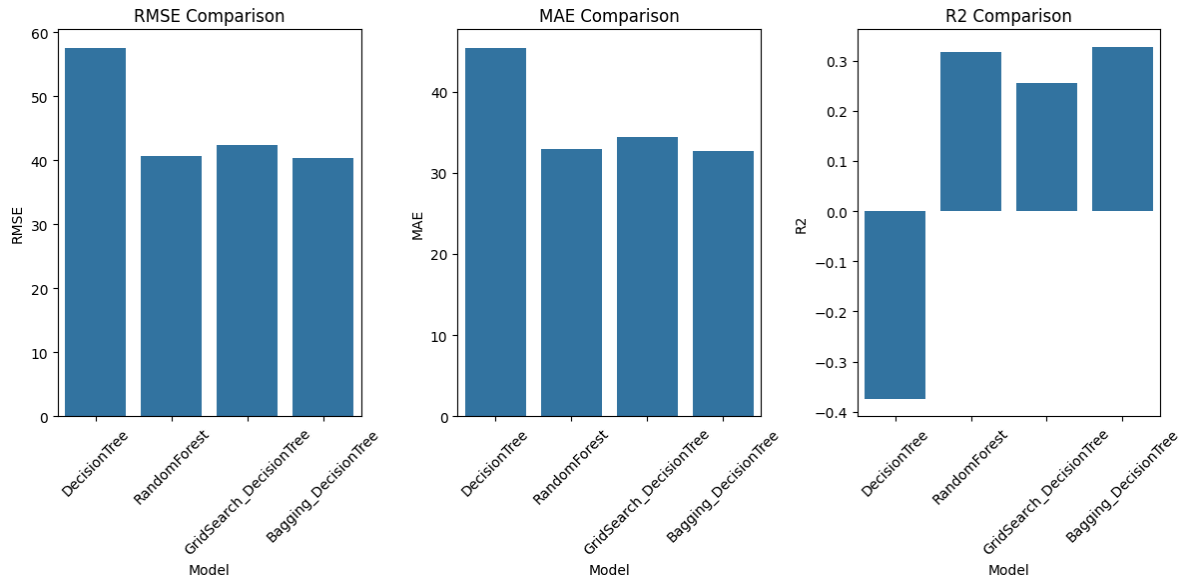


Fig. 4. Comparison of Decision Tree, Random Forest, Decision Tree (Grid Search) and Decision Tree (Bagging) for RMSE, MAE and R^2

To determine whether Decision Tree (Bagging) has a statistically significant predictive performance in comparison to the other models, we employ a paired samples t-test on 10-fold RMSE values. The comparison of Decision Tree (Bagging) with the baseline showed a statistically significant difference in RMSE scores ($t = -24.86$, $p = 0.00$), with Decision Tree (Bagging) outperforming the baseline. Similarly, Decision Tree (Bagging) performs better than Decision Tree (Grid Search) as indicated by a statistically significant difference ($t = -6.91$, $p = 0.0001$). On the other hand, there was no statistically significant difference in performance between Decision Tree (Bagging) and Random Forest ($t = -1.83$, $p = 0.10$).

To investigate how each of the models generalizes, we analyzed the model validation performance as training examples increased from 500 to 4000. The baseline Decision Tree model had the highest RMSE across all training examples, with values ranging between 55.0 to 60.0. This suggests a tendency to overfitting and instability with

varying data sizes, similar to the result reported by Cuevas-Redondo & Estevez-Bravo (2017). On the other hand, both Decision Tree (Bagging) and Random Forest maintain relatively stable RMSE values, showing consistent performance with values ranging around 40 to 42.5. However, they show limited improvement with additional training examples. There is a general downward trend in RMSE values for all the models, showing that the model performance increases as training data grows, justifying the need for optimization when modelling small datasets. The visualization curve in Fig. 5 shows the validation RMSEs across the models.

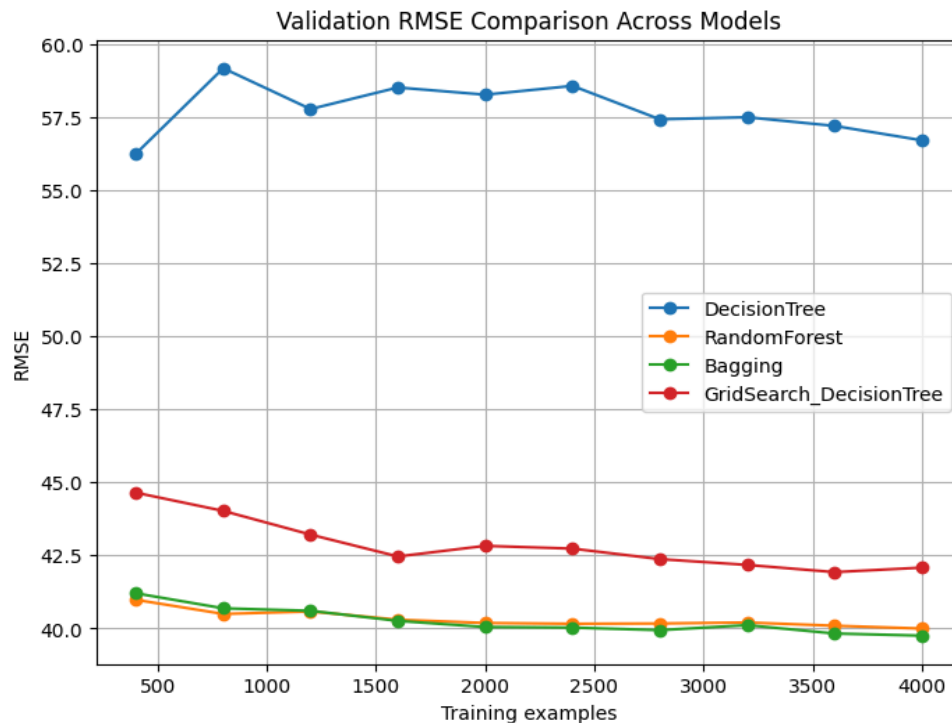


Fig. 5. Learning curves showing the performance of each of the models compared

4. Conclusion

We investigated two machine learning optimization approaches to build accurate and robust predictions. The grid Search hyperparameter tuning, which searches through a predefined set of hyperparameter values to find the best combination for a machine

learning model and the bagging ensemble method, which trains multiple models on different subsets of the training data and combines them, were compared against baseline Decision Tree and Random Forest models. The JAMB dataset downloaded from Kaggle was used for the analysis. Our results show that the baseline model has a higher MAE and RMSE as compared to the bagging ensemble models. The bagging ensemble model produced the least MAE and RMSE as compared to the grid search hyperparameter tuning. Our results indicate that bagging optimized models reduces overfitting and produce a more accurate prediction as compared to grid search and other baseline methods. Future work should look at other optimization methods like gradient descent, bayesian optimization, coordinate descent, and metaheuristic optimization.

References

- Ilemobayo, J. A., Durodola, O., Alade, O. J., Awotunde, O. J., Olanrewaju, A. T., Falana, O., Ogungbire, A., Osinuga, A., Ogunbiyi, D., Ifeanyi, A., Odezuligbo, I. E., & Edu, O. E. (2024). Hyperparameter tuning in machine learning: A comprehensive review. *Journal of Engineering Research and Reports*, 26(6), 388–395.
<https://doi.org/10.9734/jerr/2024/v26i61188>
- Aktas, H. (2025). Hyperparameter optimization for problem-based custom CNN architectures using a smart grid search method. *IEEE Access*, 13, 136374–136382.
<https://doi.org/10.1109/ACCESS.2025.3594212>
- Akuma, S., & Abakpa, H. (2021). Predicting undergraduate level students' performance using regression. *Nigerian Annals of Pure and Applied Sciences*, 4(1), 129–139.
<https://doi.org/10.46912/napas.224>

Akuma, S., & Iqbal, R. (2018). Development of relevance feedback system using regression predictive model and TF-IDF algorithm. *International Journal of Education and Management Engineering*, 8(4), 31–49. <https://doi.org/10.5815/ijeme.2018.04.04>

Alharbi, I. A., Almalki, A. J., & Zou, C. C. (2021). Hyperparameter optimization and comparison of student performance prediction algorithms. In *2021 International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 889–894). IEEE. <https://doi.org/10.1109/CSCI54926.2021.00207>

Bellaj, M., Ben Dahmane, A., Boudra, S., & Lamarti Sefian, M. (2024). Educational data mining: Employing machine learning techniques and hyperparameter optimization to improve students' academic performance. *International Journal of Online and Biomedical Engineering (IJOE)*, 20(3), 55–74. <https://doi.org/10.3991/ijoe.v20i03.46287>

Cuevas-Redondo, M. A., & Estevez-Bravo, M. (2017). Analysis techniques for the improvement and prediction of academic performance. *Universidad Complutense de Madrid*.

de Melo, G., Oliveira, S. M., Ferreira, C. C., Filho, E. P. V., Calixto, W. P., & Furriel, G. P. (2017). Evaluation techniques of machine learning in task of reprobation prediction of technical high school students. In *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)* (pp. 1–7). IEEE. <https://doi.org/10.1109/CHILECON.2017.8229739>

Foroozand, H., & Weijs, S. (2017). Entropy ensemble filter: A modified bootstrap aggregating (bagging) procedure to improve efficiency in ensemble model simulation. *Entropy*, 19(10), 520. <https://doi.org/10.3390/e19100520>

- Gul, A. (2025). A novel hybrid ensemble framework for stock price prediction: Combining bagging, boosting, dagging, and stacking. *Computational Economics*. Advance online publication. <https://doi.org/10.1007/s10614-025-10979-0>
- Hasan, M. R., Siraj, F., & Sainin, M. S. (2015). Improving ensemble decision tree performance using Adaboost and bagging. *AIP Conference Proceedings*, 030008. <https://doi.org/10.1063/1.4937027>
- Hoyos, W., & Caicedo-Castro, I. (2024). Tuning data mining models to predict secondary school academic performance. *Data*, 9(7), 86. <https://doi.org/10.3390/data9070086>
- Idowu, A. (2024). Students performance in 2024 JAMB [Dataset]. *Kaggle*. <https://www.kaggle.com/datasets/idowuadamo/students-performance-in-2024-jamb>
- Kabakchieva, D. (2012). Student performance prediction by using data mining classification algorithms. *International Journal of Computer Science and Management Research*.
- Kalita, E., Oyelere, S. S., Gaftandzhieva, S., Rajesh, K. N. V. P. S., Jagatheesaperumal, S. K., Mohamed, A., Elbarawy, Y. M., Desuky, A. S., Hussain, S., Cifci, M. A., Theodorou, P., Hilčenko, S., Hazarika, J., & Ali, T. (2025). Educational data mining: A 10-year review. *Discover Computing*, 28(1), 81. <https://doi.org/10.1007/s10791-025-09589-z>
- Menacho Chiok, C. H. (2017). Predicción del rendimiento académico aplicando técnicas de minería de datos. *Anales Científicos*, 78(1), 26. <https://doi.org/10.21704/ac.v78i1.811>
- Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research*, 11, 169–198. <https://doi.org/10.1613/jair.614>

Papadogiannis, I., Wallace, M., & Karountzou, G. (2024). Educational data mining: A foundational overview. *Encyclopedia*, 4(4), 1644–1664.

<https://doi.org/10.3390/encyclopedia4040108>

Sarmah, U., Borah, P., & Bhattacharyya, D. K. (2024). Ensemble learning methods: An empirical study. *SN Computer Science*, 5(7), 924.

<https://doi.org/10.1007/s42979-024-03252-y>

Tetko, I. V., van Deursen, R., & Godin, G. (2024). Be aware of overfitting by hyperparameter optimization! *Journal of Cheminformatics*, 16(1), 139.

<https://doi.org/10.1186/s13321-024-00934-w>

DIGITAL FORENSICS IN MOBILE DEVICES ON RESOURCES, TECHNIQUES, CHALLENGES, AND OPEN ISSUES: A SYSTEMATIC LITERATURE REVIEW

Digital Forensics in Mobile Devices on Resources, Techniques, Challenges, And Open Issues: A Systematic Literature Review

Bankole S. A., Ojeniyi J. A., Alabi I. O., Suleiman A., Noel M. D., and Isa A. O.

Department of Cyber Security Science, Federal University of Technology, Minna,

Corresponding author: bankolesamson81@st.futminna.edu.ng

Abstract

This review examines how Artificial Intelligence (AI) has transformed mobile device forensics between 2020 and 2025. It highlights AI's role in improving evidence collection, threat detection, and automating investigative processes. On one hand, AI enhances forensic capabilities by enabling detailed malware analysis, recovering encrypted data, and reconstructing timelines from fragmented information. On the other hand, it also presents new challenges as advanced threats like deepfakes and evolving malware leverage AI technologies themselves. Key challenges identified include ensuring that digital evidence remains reliable and untampered, mitigating biases that may arise from AI algorithms, and safeguarding user privacy throughout investigations. The review stresses the need to develop robust AI-based forensic frameworks alongside standardized methods for validating these tools to ensure their effectiveness and fairness. Moreover, it emphasizes the importance of international collaboration to address the global nature of cybercrime and the ethical and legal considerations involved. Ultimately, the study advocates for careful investment that supports technological progress while upholding the highest standards of privacy, fairness, and legal compliance.

Keywords: Digital Forensics, Mobile Device Forensics, Hardware Architecture, Digital Forensics Techniques, Mobile Device Techniques, Mobile Forensics, Resources.

1.0 Introduction

By 2016, about half the world was online, reshaping the digital landscape but also complicating forensic investigations since the internet wasn't designed with strong security. As cybercrime rises, experts face growing difficulties in gathering and analyzing digital evidence. To address this, forensic methods and legal frameworks must evolve. Artificial intelligence (AI) and machine learning (ML) are increasingly used to process vast and complex data quickly and accurately, improving threat detection and incident response. These technologies help manage expanding forensic databases, making investigations faster, more efficient, and better equipped to handle modern cybercrime challenges. (Dipo et al., 2024).

2.0 Related Work

Almaha et al. (2024) conducted a systematic literature review titled *"Digital Forensic Investigation on Android Devices"*. The study highlights the increasing challenges in Android mobile forensics, stressing the need for advanced tools and expertise. It emphasizes improving tools like Andriller and addressing cloud forensics to enhance investigation accuracy and efficiency. Karam & Najla (2023) conducted a comprehensive literature review titled *"Digital Forensic Tools: A Literature Review."*. This study underscores the importance of specialized forensic tools for collecting and analyzing digital evidence, comparing well-known options and newer solutions. It finds that Bayes net machine learning excels in analyzing IoT traffic. Additionally, the Stepwise Forensic Process Model helps investigators prioritize key evidence, while network packet analyzers assist in detecting unusual activity and troubleshooting. Mamoon & Khan (2013) conducted a study titled *"Exploring Static and Live Digital Forensics: Methods, Practices and Tools."*. This study highlights challenges in

collecting data during static and live digital forensics, stressing the need to preserve memory consistency for reliable evidence. It warns that some forensic tools may overwrite important data, risking investigation accuracy. The research emphasizes careful methods and tool selection to protect data integrity while effectively detecting attacks. Srinivasa & Nagalakshmi (2023) conducted a study titled *"Digital Forensics Approach for Handling Audio and Video Files"*. This study highlights the rising significance of multimedia forensics in cybercrime, focusing on audio and video evidence from devices like mobiles and CCTV. It stresses the need for specialized techniques to handle this evidence while ensuring its reliability and acceptance in court. The research advocates for standardized methods to improve the protection and management of multimedia evidence in modern investigations.

3.0 Methodology

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) The PRISMA framework, developed in 2009, helps improve transparency and rigor in systematic reviews and meta-analyses. It offers a 27-item checklist and a four-phase flow diagram to guide clear and consistent reporting. PRISMA focuses on documenting search methods, study selection, and inclusion criteria to reduce bias. The process includes identifying, screening, assessing eligibility, and final inclusion of studies. Its flow diagram visually tracks these steps, ensuring an open and verifiable review. The PRISMA guidelines improve traditional literature reviews across fields like healthcare and technology by providing flexible checklists and flow diagrams. Extensions such as PRISMA-S and PRISMA-AI support specialized research needs. Tools like flow diagrams and inclusion criteria help organize the review process and align studies with research goals. Together, they strengthen the review's rigor and support evidence-based decisions.

Table 1: Research Questions (RQ) and Rationale

S/ No	Research Area	Research Questions	Rationale
1	Resources	<i>What resources (e.g., forensic datasets, computational power, domain expertise) are most essential for developing and deploying AI-driven digital forensic models, and how do limitations in these resources affect the reliability and scalability of investigations?</i>	To understand the essentials models been employed in the development and deployment of digital forensics in mobile or smart phones.
2	Techniques	<i>Which AI/ML techniques (e.g., deep learning for file carving, NLP for log analysis, federated learning for distributed evidence collection) are employed for specific digital forensic tasks (e.g., data recovery, attribution, tampering detection), and how do their performances compare across studies?</i>	To have a deeper understanding of the kinds of AI/ML techniques that are been employed for specific digital forensics tasks such as data recovery etc. and how their performance are compared across studies.
3	Challenges	<i>What are the key technical challenges (e.g., model interpretability, adversarial attacks on forensic AI) and ethical-legal barriers (e.g., bias in training data, compliance with forensic standards) hindering the adoption of AI in digital forensics, as reported in the literature?</i>	To understand of the key technical challenges (e.g., model interpretability, adversarial attacks on forensic AI) and ethical-legal barriers (e.g., bias in training data, compliance with forensic standards) hindering the adoption of AI in digital forensics, as reported in the literature.

3.1 Information Sources and Search Strategy

This systematic review searched multiple databases, gathering 11,398 relevant publications from four major sources. Initially, 2,501 came from IEEE, 1,132 from Science

Direct, 962 from ACM Digital Library, and 479 from Springer. After removing duplicates and applying specific inclusion criteria, a final set of unique papers was identified. Most publications came from IEEE, highlighting its broad coverage of various research areas. These results form a solid base for understanding current digital forensics research and will be further analyzed to uncover key themes, trends, and gaps in the field.

3.2 Searching strings

To conduct a systematic literature review on digital forensics for mobile devices, focused on resources, techniques, challenges, and future research, specific search terms were carefully chosen based on inclusion and exclusion criteria. The search combined keywords like "mobile device forensics" with terms such as "tool selection," "hardware architecture," "challenges," and "limitations." Phrases like "systematic review" or "SLR," along with "digital forensics," "smartphone," and "techniques," were also included. Additional searches used terms related to "mobile forensics," "resources," "future research," and "open issues." Ethical factors such as "privacy & legal compliance," "consent & authorization," "data sensitivity," "bias & fairness," and "transparency & accountability" helped further narrow the results. Boolean operators like AND and OR were skillfully used to ensure a thorough and relevant collection of studies.

Table 2: Database Sources

S/NO	Database	Rationale	Supporting Reference
1	IEEE Xplore	The IEEE database is valuable for systematic literature reviews because it offers specialized, high-coverage of engineering records, making it suitable for evidence synthesis in engineering topics. Systematic reviews require comprehensive literature searches to identify all relevant records, and IEEE Xplore's coverage on engineering makes it more appropriate than multidisciplinary databases with fewer engineering records. Using specialized databases such as IEEE Xplore aligns with systematic review guidance, which advises using resources with high coverage of a specific topic.	1. Durrani et al. (2024). 2. Chigarev (2021) 3. Necula et al. (2024).
2	ScienceDirect	ScienceDirect is a valuable database for systematic literature reviews because it offers a vast collection of peer-reviewed, full-text scientific, technical, and health literature. Its multidisciplinary nature allows researchers to access scholarly research across various subject areas, making it suitable for comprehensive searches. The database's extensive coverage helps in identifying a wide range of relevant studies, which is crucial for the thorough and unbiased evidence synthesis required in a systematic review.	1. Van Dinter et al. (2021). 2. Shaheen et al. (2023).
3	ACM	The ACM Digital Library is a valuable tool for systematic literature reviews within computer science and related areas, offering a focused collection of high-quality publications. Its advanced search functionalities and comprehensive indexing enable researchers to efficiently locate relevant articles and conference proceedings, streamlining the search process. Leveraging the database's specialized content helps to synthesize evidence and establish a strong foundation for computing-related research.	1. Dillahun et al., 2017. 2. Matejka et al., 2021 3. Liu et al., 2025.

4 Springer

Scopus serves as an indispensable multidisciplinary database for systematic literature reviews, offering a rigorously curated collection of high-quality abstracts, citations, and indexed scholarly content. By encompassing a vast array of peer-reviewed journals, books, and conference proceedings, Scopus ensures that researchers have access to validated and authoritative literature, strengthening the reliability of their reviews

1. Walters, W. H. (2020).
2. Springer Nature. (2023).

Table 3: Inclusion and Exclusion Criteria

Criteria Category	Inclusion Criteria	Exclusion Criteria	Rationale
Focus	Studies that explicitly address the use of Digital forensic on mobile.	Studies that only tangentially mention Digital forensic or mobile device.	This ensured the review focuses specifically on the intersection of Digital forensic on mobile.
Study Type	Research articles, reviews, and conference reports that provide empirical data, theoretical frameworks, or practical insights.	Editorials, opinion pieces, and grey literature with inadequate information or lacking a rigorous methodology.	To prioritize studies with a robust methodology and empirical evidence.
Publication Date	Studies published within the last 5 years (2020-2025) to capture recent advancements and trends.	Studies published before 2020 to ensure the review reflects the current state of Digital forensic on mobile.	To focus on contemporary research and developments in the rapidly evolving field of Digital forensic.
Language	Studies published in English	Studies published in languages other than English.	To ensure all included studies are thoroughly assessed and understood.
Accessibility	Full-text articles that are readily accessible through academic databases used in this search.	Articles for which full text is not available or obtained within a reasonable timeframe	To ensure that all included studies can be thoroughly assessed and synthesized

3.3 Eligibility Criteria

This review looks at recent and reliable sources that specifically cover digital forensics related to mobile devices. It focuses on research that discusses forensic tools, methods, challenges, and new developments, including topics like data collection, analysis techniques, detecting anti-forensics, legal issues, and future research opportunities.

3.4 Study Selection

The study selection procedure for a systematic literature review (SLR) conducted on digital forensic follows steps outlined in Table 4.

Table 4: Study Selection Procedures

Steps	Procedures
1	Protocol Development Prior to initiating the review, the research scope, questions, search strategy, inclusion/exclusion criteria, and methods for data extraction and synthesis were established.
2	Comprehensive Literature Search Search across multiple relevant databases: Institute of Electrical and Electronics Engineers (IEEE), Science Direct, Google scholar ACM Digital Library, Science Direct, and Springer using search terms (Keywords). Result of the search in each database were filtered by relevance, date, language of publication. Pilot searches were conducted to refine search keywords based on the study objectives.
3	Study Selection (Appraisal) Studies were screen based on titles and abstracts relevant to digital forensic. Identification and selection of studies were based on the pre-defined inclusion and exclusion criteria. Since multiple databases were used, all irrelevant or duplicate articles were removed, and those that meet the eligibility requirements were retained for further assessment.
4	Full-Text Review and Quality Assessment Full text articles of selected studies were obtained. These were assessed against the inclusion and exclusion criteria. The quality and validity of each study was evaluated using the Cochrane Risk of Bias tool

or GRADE. Studies that do not meet the quality standards or inclusion criteria were excluded.

5 Data Extraction and Synthesis

Extract relevant data from the included studies using a pre-designed extraction template. This base on study design, sample characteristics, digital forensic techniques used, outcomes, and key findings. Narrative synthesis was used to draw conclusions and identify patterns or gaps in the literature.

4. Results

This segment presents a summary of the findings from the systematic review. Figure 2 illustrates the outcomes at each step of the process

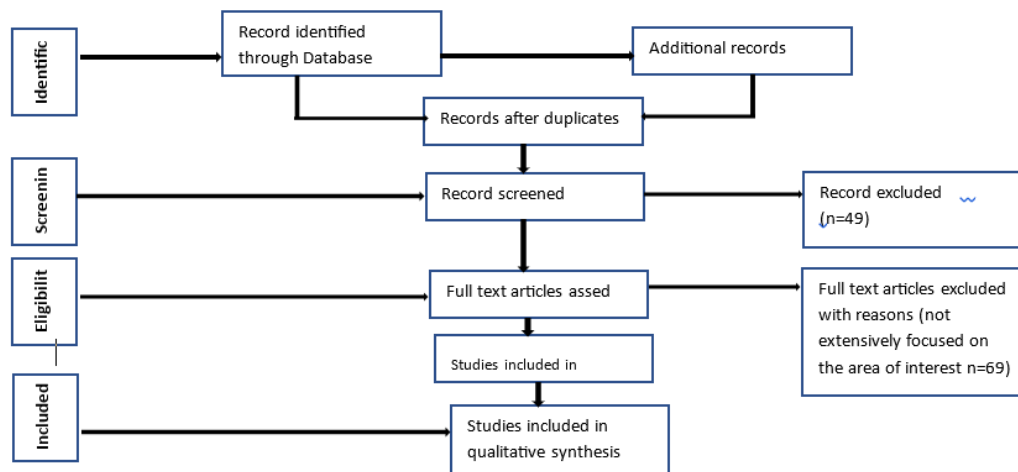


Figure 2: Result of the SLR Using PRISMA Flowchart

4.1.1 Measures of Central Tendencies and Dispersion of the Selected Articles

The measures of dispersion for the number of publications on digital forensic reveal a significant spread. The range is 2,004, indicating a substantial difference between the highest (IEEE, 2,501) and lowest (Springer, 497) values. The variance is 556,690.5, and the standard deviation is approximately 746.11 (746), further highlighting the

dispersion. These results suggest that the number of publications varies widely across the four databases.

The calculation is shown below:

Range = Maximum value - Minimum value

$$= 2,501 - 497 = 2,004$$

Variance

Variance = $\sum (xi - \mu)^2 / N$, where xi = individual data points, μ = mean, and N = number of data points

$$\begin{aligned} \text{Variance} &= (2,501 - 1,273)^2 + (1,132 - 1,273)^2 + (962 - 1,273)^2 + (497 - 1,273)^2 \\ &= (-1,228)^2 + (-141)^2 + (-311)^2 + (-776)^2 \\ &= 1,507,984 + 19,881 + 96,721 + 602,176 \\ &= 2,226,762 / 4 \\ &= 556,690.5 \end{aligned}$$

Standard Deviation

Standard Deviation = $\sqrt{\text{Variance}}$

$$= x = \sqrt{556,690.5}$$

$$= 746.11$$

To calculate the measures of central tendency:

For the mean, add up all the numbers and divide by the total number of databases (4):

$$2,501 + 1,132 + 962 + 497 = 5,092$$

$$5,092 \div 4 = 1,273$$

To calculate the median, arrange the numbers in order from smallest to largest:

497, 962, 1132, 2501

Since there are an even number of databases, the median will be the average of the two middle numbers:

$$(962 + 1,132) \div 2 = 1,047$$

The mode is the number that appears most frequently. Since IEEE has the highest number (2,501), it can be considered as the mode in a broader sense.

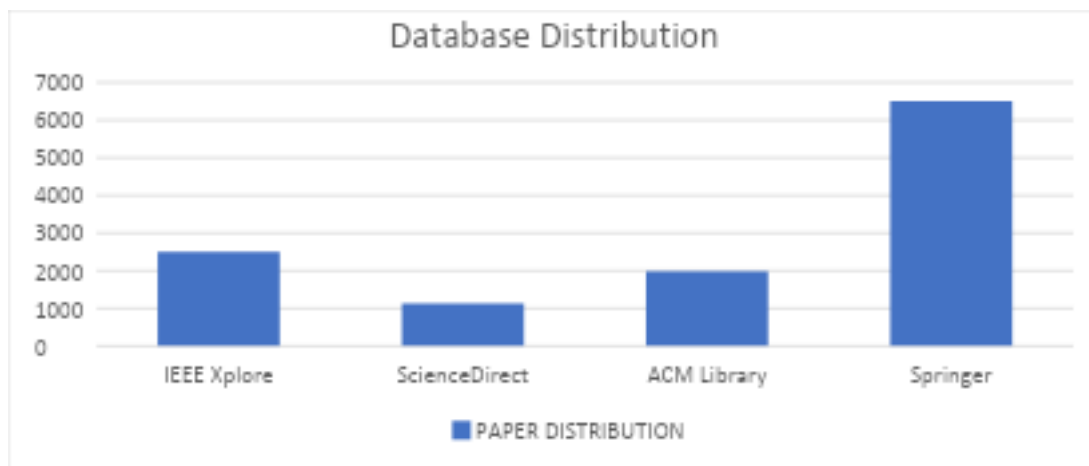


Figure 3: Database Distribution of Digital Forensics of Mobile Devices

The analysis shows that most sources come from academic journal articles, making up 43% of the total, followed by conference papers at 36%, and book chapters at 21%. This suggests a strong preference for well-reviewed and thoroughly researched material in the study.



Figure 4: Classification of Paper Type

Table 5: Selected Studies

S/no	AUTHOR	YEAR	TITLE	FOCUS	SOURCE
1	Siddharth et al.	2023	Threat modeling framework for mobile communication systems	Security threats that target or exploit mobile operator infrastructure, establishing <i>a specialized</i> approach distinct from generic IT security models.	Article history: Received 9 June 2022 Revised 12 October 2022 Accepted 1 December 2022 Available online 5 December 2022.
2	Puneet et al	2020	Enhanced Forensic Process for Improving Mobile Cloud Traceability in Cloud-Based Mobile Applications	Developing <i>traceability</i> in mobile cloud forensics through their enhanced forensic process.	ScienceDirect Procedia Computer Science 167 (2020) 907–917
3	Almaha et al	2024	A Systematic Literature Review on Digital Forensic Investigation on Android Devices	<i>An Android-centric systematic review of digital forensic investigation methodologies.</i>	ScienceDirect Procedia Computer Science 235 (2024) 1332–1352
4	Nadeem et al	2013	Systematic Literature Review on Data Carving in Digital Forensic	A <i>carving</i> -focused systematic review of digital forensic techniques.	Science direct Procedia Technology 11 (2013) 86 – 92
5	Srinivasa and Nagalakshmi	2023	Digital forensics approach for handling audio and video files	A <i>multimedia</i> -focused digital forensics approach.	Science direct Procedia Measurement: Sensors 29 (2023) 100860 Available online 7 July 2023 2665-9174/© 2023.

4.1.2 Resources for Digital Forensics of mobile devices

Mobile device forensics combines specialized hardware, software, and guidelines to effectively collect and analyze evidence. Tools like Cellebrite UFED, Magnet AXIOM, and XRY help extract data physically or logically from smartphones, while software such as Autopsy, Oxygen Forensic Detective, and MOBILedit allow detailed examination of calls, messages, apps, and cloud data. For damaged or locked devices, techniques like chip-off and JTAG are used. Standards like NIST SP 800-101 and ISO/IEC provide important reference frameworks for these processes. 27037 (Digital Evidence Collection) provides standardized methodologies for forensic investigations.

Table 6: Digital forensic resources used in mobile devices

S/N	Resources	Description	References
	Forensic Acquisition Tools		
a.	Cellebrite UFED (Universal Forensic Extraction Device)	A hardware/software solution for physical, logical, and file system extraction from smartphones (Android, iOS). Supports bypassing locks, decoding encrypted data, and cloud acquisitions.	1. Ayers et al. (2014). 2. <i>NIST SP 800-101 Rev. 1: Guidelines on Mobile Device Forensics.</i>
b.	Magnet AXIOM	A comprehensive tool for analyzing mobile backups, app data (WhatsApp, Signal), and cloud artifacts. Uses AI for timeline reconstruction.	1. Magnet Forensics (2021). <i>AXIOM Mobile Forensics Technical Guide.</i>
c.	XRY (MSAB)	Specializes in logical extraction and decoding of mobile data, including social	1. MSAB (2020). <i>XRY Digital Forensics Whitepaper.</i>

media and deleted files.

Open-Source & Analysis Tools

a. Autopsy (+ Sleuth Kit)	Open-source tool for file system analysis (SQLite, plists, APK/IPA parsing) and timeline generation.	1.Carrier (2005). <i>File System Forensic Analysis</i> .
b. Andriller	Focuses on Android forensics, extracting SMS, call logs, and app data from backups.	Andriller (2023). <i>GitHub Documentation</i> . 1.
c. Santoku Linux	A Linux distro preloaded with mobile forensic tools (ADB, Volatility, SQLite browsers) for malware analysis.	Santoku (2022). <i>Official Documentation</i> . 1.

Advanced Techniques

a. Chip-Off & JTAG Forensics	Physical extraction from NAND flash memory or via JTAG ports for locked/damaged devices.	Breeuwsma (2006). <i>Forensic Imaging of Embedded Systems</i> .
b. Cloud Forensics (iCloud/Google Backup)	Extracts synced data via APIs or authentication tokens (e.g., iCloud, Google Drive).	Quick & Choo (2014). <i>Cloud Storage Forensics</i> . 1.
c. AI-Powered Analysis	Machine learning for deepfake detection, behavioral anomalies, and encrypted data analysis.	Jessica et al. (2019). <i>Forensics Analysis on Mobile Devices: A Systematic Mapping Study</i> .

4.Standards & Guidelines

a. NIST SP 800-101 Rev. 1	Covers best practices for mobile forensics, including acquisition	NIST (2014). <i>Guidelines on Mobile Device Forensics</i> .

	methods and tool validation.	
b.ISO/IEC 27037:2012	International standard for digital evidence collection and preservation.	ISO (2012). <i>Information Technology – Security Techniques – Digital EvidenceCOLLECTIO</i> .
c. NIST CFTT (Computer Forensic Tool Testing)	Validates reliability of forensic tools (e.g., hash verification, data integrity).	NIST (2020). <i>CFTT Program Documentation</i> .
Research & Case Studies	DFRWS (Digital Forensics Research Workshop): Publishes cutting-edge research (e.g., 5G forensics, IoT device analysis).	1. DFRWS (2023). <i>Annual Conference Proceedings</i> . 2. Journal of Digital Forensics: Peer-reviewed studies on mobile malware, encryption challenges. 3. Elsevier (2022). <i>Journal of Digital Investigation</i> .

Table 7: Techniques used for digital forensics of mobile devices

S/No	Methods of digital forensics of mobile devices	Description of the mode of action	Tools	References
1.	Physical Acquisition	Creates a bit-for-bit copy of the device's physical storage (e.g., NAND flash memory).	Cellebrite UFED, XRY Physical	1. Ayers et al. (2014). <i>NIST SP 800-101 Rev. 1: Guidelines on Mobile Device Forensics</i> 2. Breeuwsma, M. (2006). <i>Forensic Imaging of Embedded Systems</i>
2.	Logical Acquisition	Extracts file system data through device APIs or backup files.	Magnet AXIOM, Oxygen Forensic Detective	1. NIST (2014). <i>Guidelines on Mobile Device Forensics</i> 2. Lessard & Kessler (2010). <i>Android Forensics: Simplifying Cell Phone Examinations</i>

3.	Chip-Off Forensics	Physically removes and reads memory chips from the device.	NAND readers, Hex editors	1. Breeuwsma (2006). <i>Forensic Imaging of Embedded Systems</i> MSAB 2. (2020). <i>Advanced Chip-Off Techniques</i>
4.	JTAG/ISP Extraction	Uses hardware interfaces to access memory directly.	RIFF Box, UFED Touch	1. NIST CFTT (2020). <i>JTAG Forensics Validation Framework</i> 2. Ayers et al. (2014). <i>NIST SP 800-101</i>
5	File System Analysis	Examines raw filesystem structures and databases.	Autopsy, Andriller	1. Carrier (2005). <i>File System Forensic Analysis</i> 2. Satish (2018). <i>iOS Forensic Analysis</i>
6.	Cloud Forensics	Extracts data from cloud backups and synced accounts.	Elcomsoft Cloud Explorer	1. Quick & Choo (2014). <i>Cloud Storage Forensics</i> 2. NIST (2021). <i>Cloud Forensic Technical Considerations</i>
7.	Live Memory Analysis	Captures volatile memory (RAM) before device shutdown.	Volatility, Belkasoft RAM Capturer	1. DFRWS (2018). <i>RAM Analysis Methodologies</i> 2. NIST (2019). <i>Memory Forensic Tools Testing</i>
8.	AI-Powered Analysis	Uses machine learning for pattern recognition and anomaly detection.	Magnet AXIOM AI, TensorFlow	1. Jessica et al. (2019). <i>Systematic Mapping Study on Mobile Forensics</i> 2. Humaira et al. (2018). <i>Validation of digital evidence.</i>
9.	Anti-Forensics Detection	Identifies and counters data hiding techniques.	Hex editors, specialized scripts	1. Garfinkel (2007). <i>Anti-Forensics: Techniques and Case Studies</i> 2. DFRWS (2020). <i>Anti-Forensic Methodologies</i>

10.	Timeline Reconstruction	Creates chronological sequences of events from device data.	Magnet AXIOM, Autopsy	1. Carrier (2005). <i>File System Forensic Analysis</i> 2. NIST (2014). <i>Mobile Device Timeline Analysis</i>
11.	App-Specific Forensics	Focuses on extracting data from specific applications.	MOBILedit, Oxygen Forensic Detective	1. Quick & Choo (2013). <i>Social Media Forensics</i> 2. NIST (2022). <i>Messaging App Forensic Analysis</i>
12.	Network Traffic Analysis	Examines device communication logs and network activity.	Wireshark, NetworkMiner	RFC 3227 (2002). <i>Guidelines for Evidence Collection</i> 2. DFRWS (2019). <i>Network Forensic Challenges</i>

4.1.4 Challenges of digital forensic of mobile device

Mobile forensics faces major hurdles due to strong encryption and security features like iOS Secure Enclave and Android FBE, making data access difficult even with advanced extraction methods. (Quick & Choo, 2014).

Table 9: challenges of digital forensic of mobile device (Summary Table)

S/No	Challenge	Mode of Action	Key References
1	Encryption	Chip-off, cloud backups	NIST (2021), Breeuwsma (2006)
2	App Fragmentation	SQLite parsing, API extraction	Quick & Choo (2014), Elcomsoft (2022)
3	Anti-Forensics	ML anomaly detection, hex analysis	Garfinkel (2007), Rosli (2025)
4	Legal Barriers	ISO 27037 compliance, warrants	ISO/IEC 27037 (2012)

4.1.8 Open Issues

The progress in AI-driven mobile forensics opens up important research opportunities. One key area is improving explainable AI to better detect unusual activity in mobile

data—like strange app behavior or encrypted messages—while keeping the process transparent enough for court use. There's also a need to develop AI tools that can piece together complex attacks that use multiple methods, such as SMS phishing, harmful apps, and cloud backups, helping investigators track down the source more quickly. Additionally, large language models (LLMs) have the potential to transform forensic training by creating realistic simulations of mobile threats, making cybersecurity professionals better prepared for real-world attacks

4.1.9 Future Directions and Resource Development

Future AI systems for threat detection will use advanced techniques like reinforcement learning and deep neural networks to operate autonomously with minimal human input. Improving the quality of training data, including using synthetic data, is essential to reduce biases and gaps. AI will also integrate multiple types of analysis—such as natural language processing, image and video recognition, and behavior tracking—to better handle complex mobile forensic challenges like detecting deepfakes through convolutional neural networks. To ensure AI-generated evidence is accepted in court, standardized evaluation frameworks like the "C-Scale" will be developed. Collaboration between humans and AI will be key, with AI managing initial data processing while humans provide contextual understanding. As criminals increasingly use AI tools for attacks like phishing and deepfakes, defenses will focus on adversarial training to strengthen AI models against such threats.

5. Conclusion

This review highlights how AI is becoming increasingly important in mobile device forensics, helping to improve investigations and tackle new challenges. Research shows that AI can automate key tasks like detecting malware in apps, analyzing

encrypted messages, and piecing together fragmented digital evidence. The review also emphasizes the importance of creating trustworthy and secure AI tools that can be confidently used in legal cases.

References

- Acharya, S., Rawat, U., & Bhatnagar, R. (2022). A comprehensive review of Android security: Threats, vulnerabilities, malware detection, and analysis. *Security and Communication Networks*. <https://doi.org/10.1155/2022/7775917>. ([ACM Digital Library](#))
- Ademu, I. O., Imafidon, C. O., & Preston, D. S. (2011). A new approach of digital forensic model for digital forensic investigation. *International Journal of Advanced Computer Science and Applications*, 2(12).
- Agarwal, A., Gupta, M., Gupta, S., & Gupta, S. C. (2011). Systematic digital forensic investigation model. *International Journal of Computer Science and Security (IJCSS)*, 5(1), 118–131.
- Ajjola, A., Zavarisky, P., & Ruhl, R. (2014). A review and comparative evaluation of forensics guidelines of NIST SP 800-101 Rev. 1:2014 and ISO/IEC 27037:2012. In *World Congress on Internet Security (WorldCIS 2014)*
- Alazab, A., Khraisat, A., & Singh, S. (2023). A review on the Internet of Things (IoT) forensics: Challenges, techniques, and evaluation of digital forensic tools.
- Al-Dhaqm, A., Ikuesan, R. A., Kebande, V. R., Abd Razak, S., Grispos, G., Choo, K. K. R., Al-Rimy, B. A. S., & Alsewari, A. A. (2021). Digital forensics subdomains: The state of the art and future directions. *IEEE Access*, 9, 152476–152502.

Alharbi, S., Weber-Jahnke, J., & Traore, I. (2011). The proactive and reactive digital forensics investigation process: A systematic literature review. In *Information Security and Assurance: ISA 2011 – Proceedings* (pp. 87–100). Springer.

Alhassan, J. K., Oguntoye, R. T., Misra, S., Adewumi, A., Maskeliūnas, R., & Damaševičius, R. (2018). Comparative evaluation of mobile forensic tools. In *Proceedings of the International Conference on Information Technology & Systems (ICITS 2018)* (pp. 105–114). Springer.

Alherbawi, N., Shukur, Z., & Sulaiman, R. (2013). Systematic literature review on data carving in digital forensic. *Procedia Technology*, 11, 86–92.

Almuqren, A., Alsuwaelim, H., Rahman, M. H., & Ibrahim, A. A. (2024). A systematic literature review on digital forensic investigation on Android devices. *Procedia Computer Science*, 235, 1332–1352.

Almehmadi, T., & Batarfi, O. (2019, May). Impact of Android phone rooting on user data integrity in mobile forensics. In *2019 2nd International Conference on Computer Applications & Information Security (ICCAIS)* (pp. 1–6). IEEE.

Al-Saleh, M. I., & Forihat, Y. (2013). Skype forensics in Android devices. *International Journal of Computers and Applications*, 38–44.

Arshad, H., Jantan, A. B., & Abiodun, O. I. (2018). Digital forensics: Review of issues in scientific validation of digital evidence. *Journal of Information Processing Systems*, 14(2), 346–376.

Aviv, I., & Ferri, U. (2023). Russian-Ukraine armed conflict: Lessons learned on the digital ecosystem. *International Journal of Critical Infrastructure Protection*, 43, 100637.

Barmpatsalou, K., Cruz, T., Monteiro, E., & Simões, P. (2018). Current and future trends in mobile device forensics: A survey. *ACM Computing Surveys (CSUR)*, 51(3), Article 54 (1–31).

Barmpatsalou, K., Cruz, T., Monteiro, E., & Simões, P. (2018). Mobile forensic data analysis: Suspicious pattern detection in mobile evidence. *IEEE Access*, 6, 59705–59727.

Baryamureeba, V., & Tushabe, F. (2004). The enhanced digital investigation process model. *Digital Investigation*

Beebe, N. L., & Clark, J. G. (2005). A hierarchical, objectives-based framework for the digital investigations process. *Digital Investigation*, 2(2), 147–167.

Breitinger, F., Hilgert, J. N., Hargreaves, C., Sheppard, J., Overdorf, R., & Scanlon, M. (2024). DFRWS EU 10-year review and future directions in digital forensic research. *Forensic Science International: Digital Investigation*, 48, 301685. <https://doi.org/10.1016/j.fsidi.2023.301685>. (DOI)

Burton, D. R. (2023). Antiviral neutralizing antibodies: From in vitro to in vivo activity. *Nature Reviews Immunology*, 23(11), 720–734. <https://doi.org/10.1038/s41577-023-00989-5>.

Camacho-Cajamarca, J. P., Campos-Argudo, K. P., Cedillo-Orellana, I. P., Coronel-Tapia, B. D., & Bermeo-Arpi, A. E. (2018). Forensics analysis on mobile devices: A systematic mapping study.

Carrier, J. G. (2005). *Gifts and commodities: Exchange and Western capitalism since 1700*. Routledge.

Dunsin, D., Ghanem, M. C., Ouazzane, K., & Vassilev, V. (2024). A comprehensive analysis of the role of artificial intelligence and machine learning in modern digital

forensics and incident response. *Forensic Science International: Digital Investigation*, 48, 301675. <https://doi.org/10.1016/j.fsidi.2023.301675>. (ScienceDirect)

Humphries, G., Nordvik, R., Manifavas, H., Cobley, P., & Sorell, M. (2021). Law enforcement educational challenges for mobile forensics. *Forensic Science International: Digital Investigation*, 38, 301129. <https://doi.org/10.1016/j.fsidi.2021.301129>.

Prakash Rao, S., Chen, H.-Y., & Aura, T. (2023). Threat modeling framework for mobile communication systems. *Computers & Security*, 125, Article 103047. <https://doi.org/10.1016/j.cose.2022.103047>. (ScienceDirect)

PREDICTION OF CRYPTOCURRENCY (BITCOIN) PRICE USING MACHINE LEARNING ALGORITHMS

Prediction of Cryptocurrency (Bitcoin) Price using Machine Learning Algorithms

Lawrence M.O., Ogbuju E., Alaebo C., and Arulogun T.

Department of Software Engineering, Miva Open University, Abuja

Corresponding Author: morolake.lawrence@miva.university

Abstract

Bitcoin is the leading cryptocurrency, having been launched in 2009. The Bitcoin market is volatile, with significant price fluctuations occurring within a limited time frame. Market speculations, emerging technologies, regulatory news and public opinion are the contributing factors for this volatility. Bitcoin price prediction has become a critical focus in financial forecasting due to its extreme unpredictability and potential economic implications. This research work evaluates the predictive performance of three machine learning (ML) algorithms: Random Forest (RF), Auto Regressive Integrated Moving Average (ARIMA), and Support Vector Machines (SVM) using previous Bitcoin price information. Assessment of approaches was based on measures such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and R-squared (R^2). RF demonstrated superior performance with a MAPE of 0.1669%, significantly outperforming ARIMA (0.6482%) and SVM (0.5548%). The outcome highlights the effectiveness of RF in identifying non-linear patterns in financial datasets. The study recommends adopting hybrid and advanced models, utilising larger datasets, and optimising model parameters to improve predictive accuracy. These recommendations aim to enhance forecasting reliability in the dynamic cryptocurrency market, paving the way for more robust financial decision-making tools.

Keywords: Bitcoin, Cryptocurrency, Time Series, ARIMA

1.0 Introduction

The digital currency known as Bitcoin is a peer-to-peer payment system that operates on a decentralised online network, meaning nobody regulates it. In 2008, Satoshi Nakamoto founded Bitcoin, the most prominent cryptocurrency (Nakamoto, 2008). The concept of a secure digital currency based on a peer-to-peer blockchain has been around for many decades. Because of blockchain technology, each piece of digital currency is encrypted and has a different digital signature, making it reliable and easier to track; hence, it has not been compromised since its introduction in 2008. It is the most popular cryptocurrency due to its exceptional price increases and volatility. Because of the nature of cryptocurrency, the supply of Bitcoin is also limited (Baur & Dimpfl, 2021). Bitcoin is unrelated to other commodities traded on the global market. Its acceptance has increased to the point where, in addition to many private businesses, El Salvador has made bitcoin its major medium of exchange and payment method (Alvarez et al., 2022). Bitcoin is now widely utilised as an investment instrument, with people trading it similarly to how they trade foreign exchange or stocks (Nguyen & Chan, 2024).

Although there wasn't much growth in the initial years following its launch, Bitcoin has experienced rapid growth since 2017. It has had periodic price increases, attracting a large number of investors. The news cycle, market events, public opinion, scarcity, demand, supply, and other standard economic factors all have an effect on cryptocurrency prices, just like they do with any other currency or investment. Due to increased demand, the Bitcoin conversion rate to the US Dollar (USD) reached an all-time high of \$64,400 USD in November 2021. Despite its enormous potential worth, daily price changes could exceed 4.61%. The subject of how Bitcoin makes money has

become increasingly popular. A nonlinear way of growing the money supply over time is called mining (Charandabi, 2021). Computers use brute force to solve mathematical equations and are rewarded with bitcoins. As a result, being able to estimate Bitcoin value is critical for ensuring a lucrative investment.

The same fundamental guidelines: "buy low, sell high". are adhered to when using Bitcoin as an investment instrument by investors. Investors who understand this notion do not invest blindly without weighing the dangers. Market technical analysis is one of the most prevalent approaches for calculating investment risks. Market technical analysis reveals the market's trend over time by analysing previous market prices. Candle graphs and market technical indicators are utilised to enhance the analysis. Even though market technical research is beneficial, interpreting the technical indications demands expertise. As a result, another automation-based solution is required.

ML, as a branch of artificial intelligence, is particularly beneficial in prediction systems. It enables the creation of prediction models that can more precisely forecast trends without the need for expert knowledge. ARIMA is a prominent ML technology for predicting future trends in the stock and FX markets. Bitcoin prices, like any other financial process, may be anticipated using the ARIMA approach. There have previously been a number of studies that looked into using an ARIMA technique to predict market value, particularly in the stock and foreign currency markets. This study proposes a Data mining approach for predicting Bitcoin's closing price using ARIMA, RF and SVM.

2.0 Review of Related Works

Hitam et al. (2019) proposed an optimised support SVM-based PSO for cryptocurrency forecasting. SVM-PSO showed the optimum results predicting the crypto coins using the day trading method. The performance accuracy of different Classifiers differs from coin to coin. Valencia et al. (2019) suggested a Sentiment analysis and ML price movement prediction. In this work, the Bitcoin blockchain network, which is typically the least useful feature for predicting the price of Bitcoin, served as the basis for the feature selection in their analysis. In the research of Lahmiri & Bekiros (2018), a deep learning chaotic neural network was employed for the prediction of the 3 most prominent traded cryptocurrencies.

Samson (2024) suggested the forecasting of Bitcoin price utilising algorithms such as DT, K-Means, LASSO, NB, and RF, examining many factors that affect the Bitcoin price. Akyildirim et al. (2020) suggested an approach to forecast cryptocurrency returns utilising ANN, RF, LR, and SVM. Among the time scale activities of daily, 1-hour and 1/2hr returns, SVM performed better. The authors recommended the incorporation of DL approaches, which improved the accuracy.

The authors in Ho et al. (2021) suggested a forecast of the Bitcoin Price model employing ML. The low, open, and high costs were utilised to predict prices of cryptocurrencies by employing a Linear regression model (LR). The investigation resulted in an accuracy of 99.8%. Although the dataset employed in this research was minute for a technique to operate on a continuous display mode, it had a high accuracy rate. In the work of Gurrib & Kamalov (2021), the forecasting of the bitcoin price was

achieved using the combination of LDA and sentiment analysis. The results achieved were compared with SVM and RF.

3.0 Methodology

The model implementation focused on applying classification techniques, including ARIMA, RF, and SVM. The input to the model consists of time-series past stock information retrieved from Yahoo Finance, an online platform offering a wide range of datasets and frameworks. The collected data is smoothed, and technical indicators are extracted from the time-series dataset. Technical indicators are factors that provide insights into future stock price behaviour. The extracted indicators are used as input features to train the ML algorithms. During the development process, a variety of tools were utilized to build and implement the system's features and functions. The models were created using Jupyter Notebook alongside the Python programming language, which provided an interactive environment for data analysis and model development. Figure 1 is the control flow diagram illustrating the sequential steps involved in cryptocurrency price prediction.

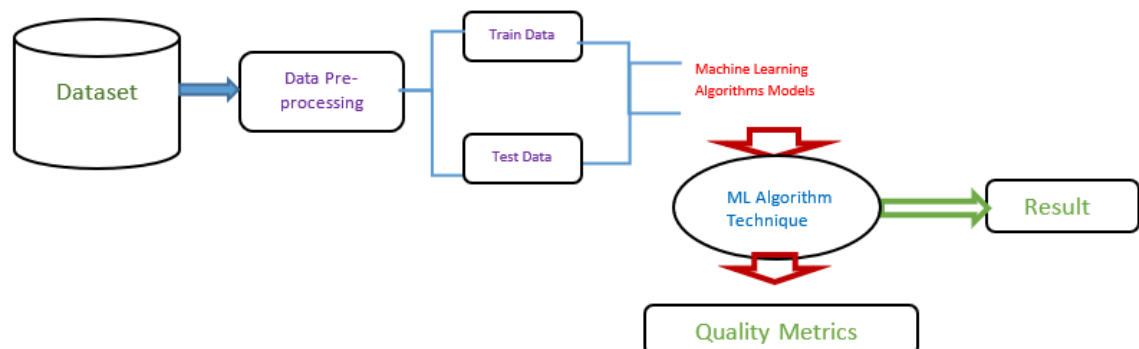


Figure 1: Control flow diagram

3.2 Preprocessing of Data

The Dataset were retrieved from the Yahoo Finance website, consisting of bitcoin prices from January 2020 to June 2022. It contains Close, High, Low, Open, and adjusted close prices on Bitcoin daily. The datasets were preprocessed to remove outliers, handle redundancy, transform the datasets, and prepare them for the selection of the most informative attributes. Some of the techniques applied to the datasets are Feature extraction and Exponential Smoothing.

Table 3 shows the sample of collected data.

Table 3: A Sample of Collected Data

	Date	Open	High	Low	Close	Adj Close	Volume	ExcelTime	ExcelCon	Diff	Timestamp
0	1/1/2020	7194.892090	7254.330566	7174.944336	7200.174316	7200.174316	1.856566e+10	43831	25569	18262	1577836800
1	1/2/2020	7202.551270	7212.155273	6935.270020	6985.470215	6985.470215	2.080208e+10	43832	25569	18263	1577923200
2	1/3/2020	6984.428711	7413.715332	6914.996094	7344.884277	7344.884277	2.811148e+10	43833	25569	18264	1578009600
3	1/4/2020	7345.375488	7427.385742	7309.514160	7410.656738	7410.656738	1.844427e+10	43834	25569	18265	1578096000
4	1/5/2020	7410.451660	7544.497070	7400.535645	7411.317383	7411.317383	1.972507e+10	43835	25569	18266	1578182400
5	1/6/2020	7410.452148	7781.867188	7409.292969	7769.219238	7769.219238	2.327626e+10	43836	25569	18267	1578268800
6	1/7/2020	7768.682129	8178.215820	7768.227539	8163.692383	8163.692383	2.876729e+10	43837	25569	18268	1578355200
7	1/8/2020	8161.935547	8396.738281	7956.774414	8079.862793	8079.862793	3.167256e+10	43838	25569	18269	1578441600
8	1/9/2020	8082.295898	8082.295898	7842.403809	7879.071289	7879.071289	2.404599e+10	43839	25569	18270	1578528000
9	1/10/2020	7878.307617	8166.554199	7726.774902	8166.554199	8166.554199	2.871458e+10	43840	25569	18271	1578614400

3.1.1 Exponential Smoothing

Firstly, an exponential smoothing is applied to the time series dataset. It is a general method for smoothing the dataset by employing the exponential window function. It works by exponentially reducing the weights of earlier observations while increasing the weights of more recent observations. The result of the exponential smoothing process is typically written as $\{S_t\}$, which might be considered as the best indication of

what the next value of x will be. The raw data series is commonly represented a, $\{x_t\}$ starting at the time. The following formulas provide the most basic version of exponential smoothing when the series of observations starts at time $t = 0$

$$S_0 = X_0 \quad (1)$$

$$S_t = \alpha Y_t + (1 - \alpha) * S_{t-1}, t > 0 \quad (2)$$

$$S_t = \alpha \text{ where } \alpha \text{ is a smoothing factor, and } 0 < \alpha < 1$$

3.1.2 Dimensionality Reduction

The act of converting data from a high-dimensional space to a low-dimensional space while preserving some informative features of the initial data is called Dimension Reduction. New features are created from the existing datasets, hence reducing the number of features in the dataset. The dimension reduction technique employed in this work is Feature extraction, where the most relevant features are extracted for the model to predict from.

Feature Extraction

Feature extraction reduces the dimension of a large amount of data by breaking it up into smaller groups for processing. Its goal is to reduce the number of features in a data set by repurposing existing data to create new data. The original set of features should then be able to summarise the majority of the information in the new, reduced set of features. Technical Indicators are essential metrics that are computed from time series stock data and are used to forecast the direction of the financial markets. They are instruments that investors commonly employ to look for bearish or bullish signs. Technical indicators employed for this research are explained as follows:-

i. Relative Strength Index

The Relative Strength Index (RSI) is a popular momentum indicator for determining whether a stock is overbought or oversold. When demand unjustifiably pushes the price of a stock upwards, it is considered to be overbought. This is usually viewed as an indication that the stock is overvalued and is likely to fall in price. When a stock's price falls drastically below its true value, it is considered to be oversold. As a result of panic selling, this is the outcome. The RSI scale runs from 0 to 100, and a reading of more than 70 indicates that the stock is overbought, while a reading of less than 30 indicates that the stock is oversold.

RSI is computed as:

$$RSI = 100 - \frac{100}{1+RS} \quad (3)$$

$$\text{Where } RS = \frac{\text{Average gain over the past 14 days}}{\text{Average loss over past 14 days}}$$

ii. Stochastic Oscillator

A momentum indicator called the stochastic oscillator contrasts a security's closing price with a range of its prices over a given time frame. It tracks the price's velocity or momentum. In most cases, momentum shifts before the price shifts. It calculates the level of the closing price concerning the low-high range throughout time.

$$\%K = (C - L_{14}) / (H_{14} - L_{14}) * 100 \quad (4)$$

where,

C = Current Closing Price

L₁₄ = Lowest Low over the past 14 days

H₁₄ = Highest High over the past 14 days

iii. Williams %R or Williams R

Williams %R, or simply %R, is a metric that displays the current closing price in proportion to the previous N-day high and low. It has a range of -100 to 0. It indicates a sell signal when its value is greater than -20, and a buy signal when its value is less than -80.

It is computed as follows:-

$$Williams \% R = \frac{Highest\ High - Lowest\ Low}{Highest\ High - Close} \quad (5)$$

Where:

Highest high = the highest price within the 14-day lookback period

Close = Current Closing price

Lowest Low = the cheapest price during the 14-day lookback period

iv. PROC (Price Rate of Change)

Based on momentum, the PROC is a technical indicator that determines the percentage change in price between the present and the previous prices. It computes the most recent price change in relation to the price n days ago.

$$P_{ROC}(t) = \frac{C(t) - C(t-n)}{C(t-n)} \quad (6)$$

Where:-

$C(t)$ = the most current period closing price

$C(t - n)$ = the closing price n times before the latest time frame

3.3 Prediction Algorithms

The algorithms were employed in this research: ARIMA, RF, and SVM.

1. ARIMA

A simple stochastic time series model, ARIMA, can be used to forecast and train future time points. In the stock and foreign exchange markets, ARIMA prediction systems are frequently used. To make an autoregressive integrated moving average model stationary, the data are differenced. A technique for turning a non-stationary time series into a stationary one is called differencing. The data's consistency throughout time is shown via a stationarity model. The time series is deemed non-stationary and exhibits seasonality if there is a pattern. The purpose of differencing is to remove any trends or seasonal structure that are present in the majority of market and economic data. Time series data for the variable of interest are computed to perform the ARIMA prediction. ARIMA prediction is accomplished by entering time series data for the variable of interest. Jupyter Notebook determines the number of delays or amount of differencing to apply to the data and verifies for stationarity (meaning the mean and variance are constant over time). Figure 2 shows a graphical representation of the application model of ARIMA in which the time series dataset is passed into the ARIMA model created, and then the IDE determines seasonality and differencing.

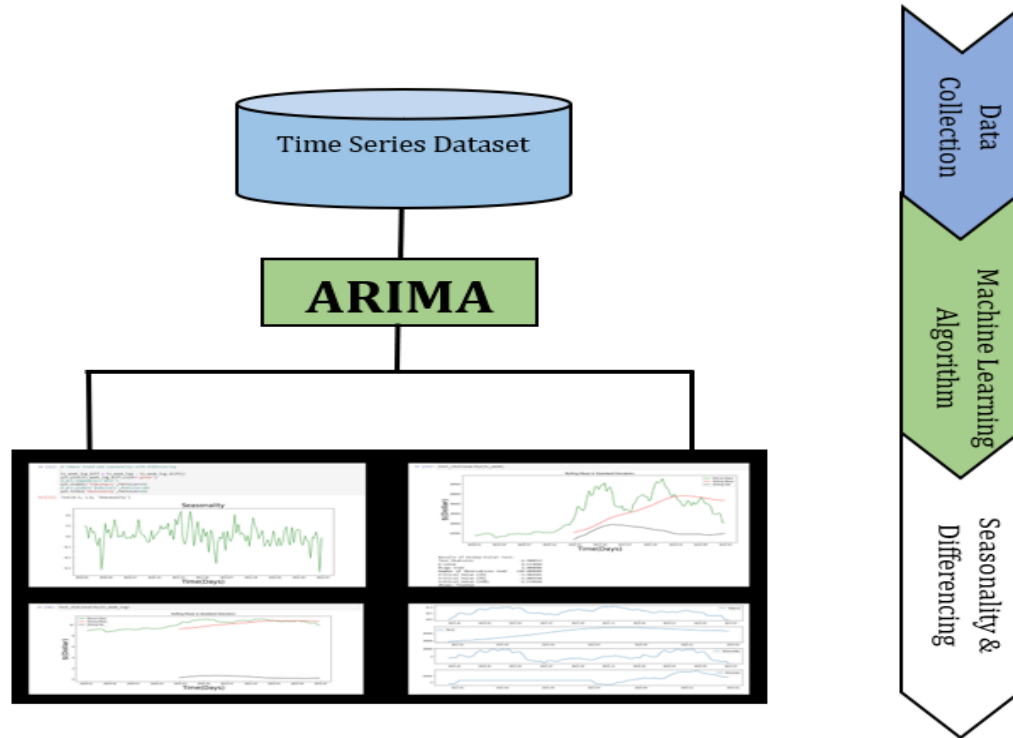


Figure 2: Application of the ARIMA model (Namin, 2018)

Table 1: Algorithm for ARIMA (Namin, 2018)

Input	Training dataset T
Output	A class of testing datasets
Steps	1. Use the data to create a time series graph. 2. Adjust the data so that it is stationary on the mean 3. To make the data variance-independent, log transform it. 4. Apply a difference log transform to the data to make the variance and mean stationary 5. Use decomposition to eliminate trend and seasonality 6. Use the ARIMA model to forecast the price

2. RF

RF is a decision tree-based supervised ML technique. This means it can be used to forecast time series data, provided the time series dataset is first converted into a supervised learning problem. It also necessitates the use of a specific model evaluation technique known as walk-forward validation, as k-fold cross-validation would yield optimistically biased results. This algorithm is used to anticipate behaviour and outcomes in a variety of industries, such as banking and e-commerce. Our price prediction algorithm would be based on the patterns generated by these decision trees. The more data we can feed the RF algorithm, the more chances it has of discovering new patterns and validating old ones. It will then take the average of each tree's forecasts to build a more credible prediction. Figure 3 shows a graphical representation of how the RF Classifier works, how decisions are made from the obtained dataset, and the final prediction made.

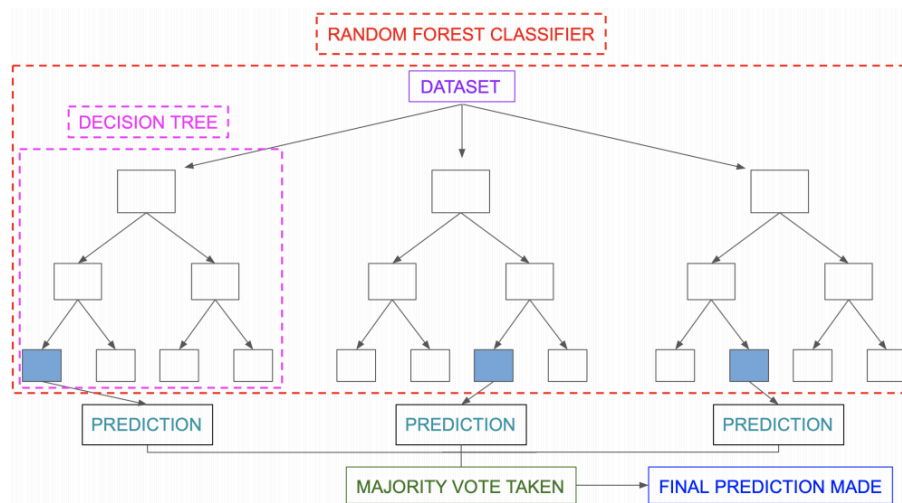


Figure 3: Random Forest Classifier works

The RF is formed in two phases: the first is to combine N decision trees to build the RF, and the second is to make predictions for each tree created in the first phase. Table 2 shows the list of the steps involved:

Table 2: Steps for Random Forest Algorithm

Steps	1. Choose K data points at random from the training set. 2. Construct the DT corresponding to the chosen data points 3. Select the number N for DT that you want to create. 4. Steps 1 & 2 should be repeated. 5. Determine each DT's predictions for the new data points, then allocate them to the category with the most votes.

3 SVM

SVM is a supervised ML approach that can perform regression, classification, and even outlier detection. SVMs have lately been employed in a variety of applications, including data mining, classification, regression, and time series forecasting. The concept of SVM is straightforward: The method draws a line or a hyperplane to divide the data into classes. An SVM establishes a decision boundary between classes by employing support vectors, which are instances of training data. To attain the most remarkable hyperplane separation between instances of the different classes, this is done (Lawrence, 2024). Figure 4 shows how a support vector machine works and how it categorises data points even when they are not otherwise linearly separable by mapping the data to a high-dimensional feature space.

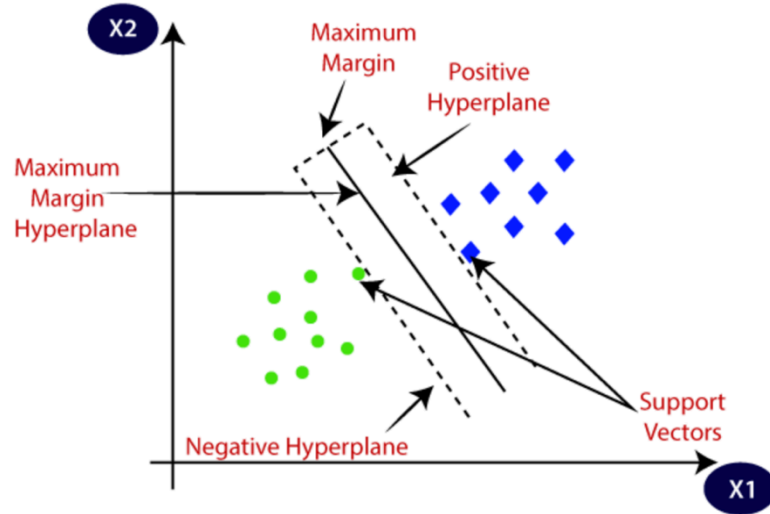


Figure 4. Support Vector Machines

3.4 Performance Metrics

1. Mean Absolute Error (MAE)

The average of the absolute differences between the dataset's actual and anticipated values is known as the MAE. It calculates the dataset's average residual. Stated differently, MAE is a metric that quantifies the variation in error between two observations that exhibit identical behaviour. It is computed as follows:-

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (7)$$

where,

$\hat{Y}$ = Predicted Value of y

Y = True value of y

2. Mean Squared Error (MSE)

The average squared difference between the dataset's initial and anticipated values is known as the MSE. It calculates the residuals' variance.

The formula for calculating MSE is given as:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (8)$$

3.4.3 Root Mean Squared Error (RMSE)

The square root of MSE is RMSE. It calculates the residuals' standard deviation. Its formula is given as: $RMSE = \sqrt{MSE}$

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (9)$$

3. Mean Absolute Percentage Error (MAPE)

This is a measure for computing the prediction accuracy.

$$\frac{100}{n} \sum_i \frac{|Y_i - \hat{Y}_i|}{Y_i} \quad (10)$$

4.0 Results and Discussion

The dataset was collected from Yahoo Finance (a media property part of the Yahoo! Network that delivers stock quotations, press releases, financial reports, and original material, as well as financial news, data, and opinion) website, which provides historical Bitcoin prices from January 2020 to June 2022. The dataset contains Bitcoin's Open, High, Low, Close, and Adjusted Close (the closing price after applying adjustments for any applicable splits and dividend payouts) prices on a daily interval. Figure 4 is a graphical chart that displays the daily closing price of Bitcoin (in USD) over

5 months beginning January 2022, using Japanese candlesticks. Figures 5, 6, and 7 show the time series datasets and prediction in ARIMA.

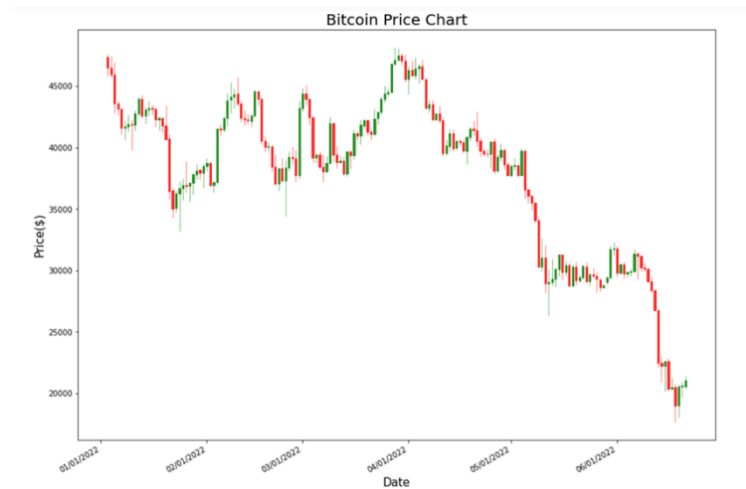


Figure 4 is a graphical chart that displays the daily closing price of Bitcoin

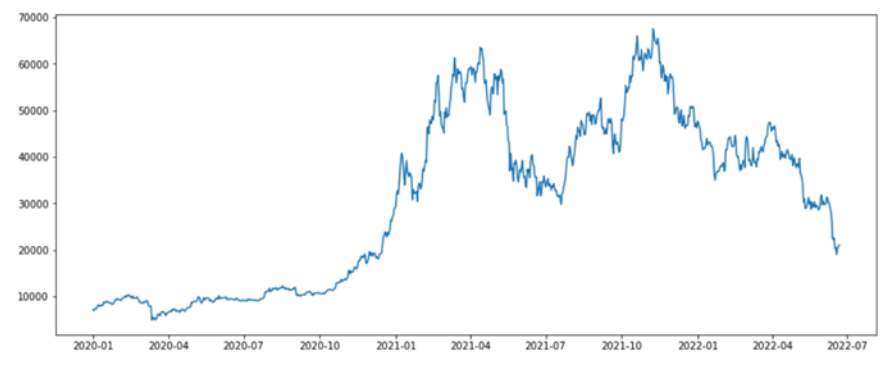


Figure 5. Visualising the Time Series Data Set

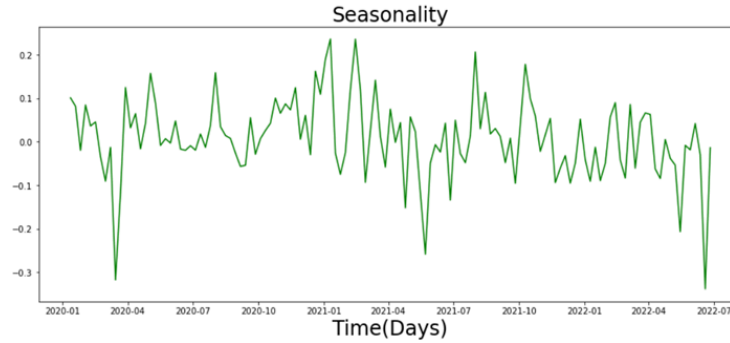


Figure 6. Seasonality in ARIMA

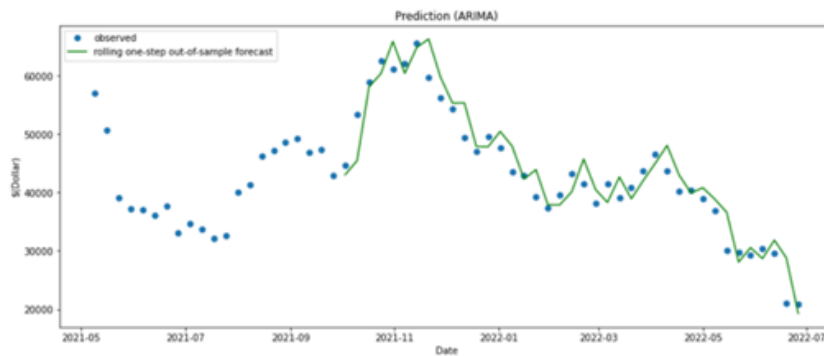


Figure 7. Prediction Result using ARIMA

4.1 Performance Evaluation

The prediction performance evaluation metrics provide a summary of the forecast model's capacity, skill, and capability in making the forecasts. This research's models utilized 5 performance indicators, which include: MAE, MAPE, MSE, RMSE, and R^2 , but the focus for this project is MAPE because it is very common and easy to read (interpret). Table 1 shows the Performance Evaluation of ARIMA, Random Forest, and SVM algorithms in terms of MAPE, MAE, MSE, RSME, and R-squared. This table shows that Random Forest had a good MAPE score of 0.1669% compared to ARIMA (0.6482) and SVM (0.5548%).

Table 4: Performance Evaluation

Performance Measures	ARIMA	Random Forest	SVM
MAPE	0.6482%	0.1669%	0.5548%
MAE	0.068343	5474.0113	19118.46
MSE	0.007819	66799585.197	281231905.26
RMSE	0.0884	0.4085	16787.8499
R Squared	0.8918	0.7837	0.1496

Table 5 presents a comparison of algorithms used by relevant and related authors and researchers, along with their results (evaluated using the mean absolute percentage error for performance assessment).

Table 5 Comparison of other models in the literature

Authors	Algorithm/Method	Results
Hamayel <i>et al.</i> (2021)	LSTM and bi-LSTM, GRU	1.1234%, 5.990%, 0.2454%, respectively (MAPE)
Hansun <i>et al.</i> (2022)	LSTM and bi-LSTM, GRU	0.0997, 0.0725, 0.0698% respectively (MAPE)
Aygün <i>et al.</i> (2020)	One-layer ANN, Three-layer ANN, CNN, RNN	0.07, 0.22, 0.40, 0.25% respectively (MAPE)
Derbentsev <i>et al.</i> (2021)	Random Forest and Stochastic Gradient Boosting Machine	2.61% and 2.31% respectively (MAPE), 305.95 and 263.34 (RMSE)

5.0 Conclusion

Three machine learning models were used in this study to predict the closing price of the Bitcoin cryptocurrency to a certain extent and with some degree of accuracy: ARIMA, random forest classification model, and support vector machine classification

were the algorithms used. The results of this ARIMA suggest that, given bitcoin's price sensitivity to sharp rises and falls, predictions of its price based on its closing price history may provide high MSE values. It is noted that numerous decisions must be made in Machine learning algorithms regarding the appropriate techniques, data segmentation, cross-validation, feature collection, etc. There is no certainty that we are employing the best techniques because the primary objective of this study is not to thoroughly and accurately measure different forecasting and trading techniques. Instead, we're just checking to see how machine learning can contribute to the prediction of Bitcoin price, as it is widely affected by a lot of economic activities. Perhaps as time passes and technological, artificial intelligence, machine learning, and national economies become more stable, it will be possible to anticipate the price of bitcoin accurately.

References

- Akyildirim, E., Corbet, S., Sensoy, A., & Yarovaya, L. (2020). The impact of blockchain related name changes on corporate performance. *Journal of Corporate Finance*, 65, 101759. <https://doi.org/10.1016/j.jcorpfin.2020.101759>
- Alvarez, F., Argente, D., & Van Patten, D. (2022). Are cryptocurrencies currencies? Bitcoin as legal tender in El Salvador. *National Bureau of Economic Research Working Paper Series*. <https://doi.org/10.3386/w29968>
- Aygun, B., & Gunay, E. K. (2021). Günlük Bitcoin değerini tahmin etmek için istatistiksel ve makine öğrenimi algoritmalarının karşılaştırılması. *European Journal of Science and Technology*. <https://doi.org/10.31590/ejosat.822153>

Baur, D. G., & Dimpfl, T. (2021). The volatility of Bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics*, 61, 2663–2683.

<https://doi.org/10.1007/s00181-020-01990-5>

Charandabi, S. E., & Kamyar, K. (2021). Prediction of cryptocurrency price index using artificial neural networks: A survey of the literature. *European Journal of Business and Management Research*, 6(6), 17–20.

<https://doi.org/10.24018/ejbmr.2021.6.6.1138>

Derbentsev, V., Datsenko, N., Babenko, V., & Pushko. (2020). Forecasting cryptocurrency prices using ensembles-based machine learning approach. In *2020 IEEE International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)* (pp. 707–712). IEEE.

<https://doi.org/10.1109/PICST51311.2020.9468090>

Gurrib, I., & Kamalov, F. (2021). Predicting Bitcoin price movements using sentiment analysis: A machine learning approach. *Studies in Economics and Finance*, 39(3), 347–364. <https://doi.org/10.1108/sef-07-2021-0293>

Hamayel, M. J., & Owda, A. Y. (2021). A novel cryptocurrency price prediction model using GRU, LSTM and bi-LSTM machine learning algorithms. *AI*, 2(4), 477–496. <https://doi.org/10.3390/ai2040030>

Hansun, S., Wicaksana, A., & Khaliq, A. Q. M. (2022). Multivariate cryptocurrency prediction: Comparative analysis of three recurrent neural networks approaches. *Journal of Big Data*, 9(1).

<https://doi.org/10.1186/s40537-022-00601-7>

Hitam, N. A., Ismail, A. R., & Saeed, F. (2019). An optimized support vector machine (SVM) based on particle swarm optimization (PSO) for cryptocurrency

forecasting. *Procedia Computer Science*, 163, 427–433.

<https://doi.org/10.1016/j.procs.2019.12.125>

Ho, A., Vatambeti, R., & Ravichandran, S. K. (2021). Bitcoin price prediction using machine learning and artificial neural network model. *Indian Journal of Science and Technology*, 14(27), 2300–2308. <https://doi.org/10.17485/ijst/v14i27.878>

Lahmiri, S., & Bekiros, S. (2018). Cryptocurrency forecasting with deep learning chaotic neural networks. *Chaos, Solitons & Fractals*, 118, 35–40. <https://doi.org/10.1016/j.chaos.2018.11.014>

Lawrence, M. O. (2024). SqueezeNet deep learning model for magnetic resonance imaging brain tumor detection. *Faculty of Natural and Applied Sciences Journal of Scientific Innovations*, 6(1), 72–83. <https://fnasjournals.com/index.php/FNAS-JSI/article/view/559>

Namin, A. S. (2018). Forecasting economics and financial time series: ARIMA vs. LSTM. *arXiv*. <https://arxiv.org/abs/1803.06386>

Nguyen, D. T., & Chan, K. C. (2024). Cryptocurrency trading: A systematic mapping study. *International Journal of Information Management Data Insights*, 4(2), 100240. <https://doi.org/10.1016/j.jjime.2024.100240>

Samson, T. K. (2024). Comparative analysis of machine learning algorithms for daily cryptocurrency price prediction. *Information Dynamics and Applications*, 3(1), 64–76. <https://doi.org/10.56578/ida030105>

ARTIFICIAL INTELLIGENCE IN HUMAN RESOURCE MANAGEMENT: CHALLENGES AND FUTURE DIRECTIONS

Artificial Intelligence in Human Resource Management: Challenges and Future Directions

¹Obanor I. Enoch, ²Agboola G. Mayowa, ²Nnodum C. Jessica, ²Onumajuru I. Richard,

²Umeh U. David, ¹Olabode D. Stephen

¹Department of Business Management, Miva Open University

²Department of Business Management, Covenant University

Corresponding author: enoch.obanor@miva.edu.ng

Abstract

Artificial Intelligence (AI) is currently transforming the decision-making processes in the project and human resource management by enhancing the standard Decision Support Systems (DSS). DSS powered by AI can measure the elements of efficiency and accuracy by increasing their automation, analysis in real-time, prediction, and advanced risk screening. Such systems enable managers to maximize their resource allocation efforts, minimize the execution of work, and reduce human error that eventually results in enhancing project results and workforce performance. However, things are not entirely peachy about the implementation of AI-based DSS. There are barriers, e.g., the data quality, ethics, interfacing the system with the existing programs, and organizational reluctance to change. This mini-review identifies the potential and possible positive benefits, challenges, and the future of AI in DSS. This review will shed some light on how AI can further be used to come up with smarter frameworks of the decision system in various facets of management by analysing the latest developments and practical obstacles.

Keywords: Artificial Intelligence, Decision Support Systems, Human Resource Management, Predictive Analytics

1. Introduction

Artificial intelligence (AI) is not only changing industries but is making them more efficient by improving decision-making procedures, automating repetitive activities, and making all the processes more efficient (Mathew et al., 2023). One of the areas that has benefited greatly because of the advent of AI is project management and human resource management, especially through the integration of AI in Decision Support Systems (DSS) (Waqar, 2024). The long-used traditional DSS, serving to accumulate and compile information that can be used in making more profitable decisions, is now complemented by AI, which adds much more functionality and contributes to making DSS much more useful (Gupta et al., 2022).

The use of AI-based DSS puts real-time data analysis, predictive analytics, and refined risk assessment in the limelight, and managers can make informed decisions in a short time. As another example, in project management, AI solutions have the potential to automate the project processes by streamlining the budget support, scheduling, and resource allocation, enabling the project manager to conduct the work on a more strategic level (Tanim and Ahmad, 2025). Likewise, AI can support talent acquisition, performance monitoring, and workforce analysis in human resource management and enable organizations to optimise their human capital (Bashynska et al., 2023). These developments show that AI has the opportunity to improve accuracy and efficiency in the world of management (Gupta et al., 2022).

Nevertheless, as well as providing such advantages, incorporating AI into DSS, it comes along with peculiar challenges. The purpose of this mini-review is to find out what opportunities and limitations AI in DSS has concerning the human resource management environment. This review will give information that will guide the use of AI-enabled DSS to promote efficiency and reduce the challenges of implementation

and the emergence of smarter, agile decision-making frameworks with the help of literature and practice. This knowledge can be instrumental in managing the consequences of such dynamics as organizations tap into the power of AI and figure out how to streamline its intricacies.

2. Key Opportunities of AI in Human Resource Management

The use of AI in human resources (HR) is deeply beneficial; however, it comes with a variety of issues. On the bright side, AI has the potential to automate numerous HR tasks. As a case in point, Nawaz et al. (2024) observe that it is possible to entrust the time- and labor-intensive duties related to HR management, including applicant filtering and scheduling to AI to streamline overall quality and effectiveness of HR operations. In another similarity, Abdelhay et al. (2025) reveal that due to the significant level of bias elimination and the enhancement of the level of accuracy, a generative AI tool can be used to short-list the candidate faster, more comprehensively, and with fewer inclinations to errors and biases. Practically, research indicates that AI has the capability of automating the development of job descriptions, the screening of resumes as well as communicating with candidates, which saves the HR professionals a lot of time and allows them to concentrate on their strategic processes (Nawaz et al., 2024; Abdelhay et al., 2025). The workforce planning and individual development can be further enhanced with HR analytics that uses data to generate a competitive edge in organisations.

Some of the major AI opportunities in HR are:-

Productivity and efficiency: Automation of routine activities (e.g. sourcing, resume screening) can enhance the productivity of HR to a large extent. It is identified in the research that the results of AI adoption are a set of time-saving, cost-efficient outcomes with augmented accuracy and workload in general (Nawaz et al., 2024). As

an example, Dutta et al. (2025) adds that AI-based tools to organize schedules and interviews have already displaced most administrative tasks, allowing recruiters to concentrate on advising of the highest value. This will increase productivity and efficiency at the workplace.

Reduction of cost: AI would help minimize cost related to the recruitment process and administration through the minimization of manual operations and speeding it up. Nawaz et al. (2024) directly name time and cost savings as some of the advantages of AI by conducting surveys among HR practitioners on an empirical basis. According to these surveys, the costs associated with each hire become lower in when AI is involved in screening and interviewing (Nawaz et al., 2024).

Data-driven decision-making: AI also enables HR professionals to leverage large datasets for forecasting and strategic planning. Advanced analytics and machine learning can predict skills gaps, tailor training, and identify top talent more objectively (Dutta et al., 2025). According to Abdelhay et al. (2025), generative AI's "algorithmic objectivity" yields more consistent evaluation of candidates, potentially enhancing the quality and diversity of hires. Overall, AI's insights help move HR from intuition-based to evidence-based decision processes, improving organisational agility and competitiveness.

3. AI Challenges in Human Resource Management

In spite of all the advantages of AI in HR, there are some issues waiting to be solved. One of the priorities is ethical and bias. Artificial intelligence systems can create or even increase current forms of prejudice when provided with biased information (Chen, 2023). Chen (2023) concludes that the AI recruitment can create a

discriminatory effect (based on gender, race, etc.) due to its algorithmic bias unless it is properly circumvented. Likewise, another similarity in the available research on ethical aspects of using AI in HRM is the emphasis of Potluri and Serikbay (2025), who point to the ethical considerations of AI regarding employee data privacy and algorithmic bias as the key impediments to its adoption. Thoughtless use of AI systems carries a chance of unjust treatment of the applicants and may destroy trust in the HR. Specifically, any kind of “black box” algorithm is hard to hold accountable and lacks any explainability, which means that employees and candidates might see AI decisions as arbitrary (Patil, 2025; Obanor et al., 2025).

Almost connected with it is the problem of data privacy. The work of HR services touches upon information that is related to sensitive personal data, and therefore, ethical regulations and laws regarding data protection have to be addressed by AI. Potluri and Serikbay (2025) emphasize that when HR data are not regulated clearly and governed properly, such a phenomenon poses legal threats. As an example, AI recruiting technologies that analyse candidate information will be required to be respectful to various frameworks such as the General Data Protection Regulation (GDPR), because non-compliance may lead to fines and the slackening of employee trust (Boelens, 2024). Employees are afraid of predictive tracking even without control: in case employees realize that their own information (e.g. performance indicators, personal profiles) is abused, they can be opposed to AI devices (Patil, 2025). It has been observed that the fear of privacy and prejudice might deter workers from using AI, limiting the quality of data and reducing the effectiveness of systems (Potluri and Serikbay, 2025; Boelens, 2024). In this way, data security and clear data usage are essential issues.

Resistance to change in organizations is another big challenge. Several employees and leaders fear the replacement of human opinion or employment via AI (Park et al., 2021). According to Potluri and Serikbay (2025), cultural resistance represents a significant deterrent with regard to implementing AI: organisations state that workers as well as management tend to mistrust the AI-new tools. Uncertainty is present with change, and unless the HR initiative has been adequately communicated and trained, it will be resisted. This relates well to the skills gap issue. The implementation of AI changes the professional skills of the HR personnel by requiring changes in technical and digital skills. In line with this, Potluri and Serikbay (2025) state that most HR professionals are not yet experts in the application of AI tools and other forms of analytics, and this poses a bottleneck towards successful utilisation. These gaps in skills are both technical (e.g., the ability to understand algorithms) and conceptual (being able to relate the output of AI to strategy), and soft (communicating the value of AI to other team members). The HR teams can roll out AI incorrectly or never put it into use in the absence of education and upskilling (Cai et al., 2024).

There are also challenges brought about by infrastructure and resource constraints. The current legacy HR systems and data silos may complicate deployment of AI (Barach, 2025). There is also the need in considerable time and capital when upgrading or incorporating infrastructure, particularly in smaller organisations (Jatob, 2023). According to Potluri and Serikbay (2025), the use of old software can lead to incompatibility with modern AI platforms; moreover, the migration of data can disrupt its work. Inadequate computer resources, low quality data or disjointed IT serve as a drag to uptake. Moreover, minor companies might have insufficient IT governance to inspect AI products, which leads to the intensification of the threat of uncontrolled algorithm mistakes (Jatobá et al., 2023).

Finally, the legal and regulatory environment is still evolving. While laws like GDPR address some privacy aspects, many countries have no specific rules for AI in hiring or HR decisions (Boelens, 2024). This regulatory uncertainty itself is a challenge because organisations fear non-compliance and liability. Potluri and Serikbay (2025) explain that without clear AI regulations, HR divisions face the risk of inadvertently violating data protection laws and anti-discrimination statutes. In response, jurisdictions like the EU are introducing comprehensive AI laws (e.g. the EU AI Act) that classify HR analytics and recruitment algorithms as high-risk, requiring strict oversight (Malin et al., 2025). This creates pressure on employers to prepare formal compliance strategies even as rules continue to crystallise.

Overall, organisations adopting AI in HR must balance these concerns and benefits, ensuring ethical use and robust oversight while leveraging technology to achieve greater efficiency and strategic insight (Boelens, 2024; Malin et al., 2025).

4. Emerging Trends and Future Directions

In the future, AI in HR is undergoing ever-increasing changes supported by new technologies and regulatory change. The generative AI is a growing trend. Human resources (HR) are currently adopting large language models (e.g. ChatGPT) across various tasks, such as the creation of a job advert, providing information to candidates, creating training materials, even simulated interviews (Maree and Shehada, 2024). According to recent studies, generative AI considerably increases recruitment efficiency through automation of the resume analysis and communication. As an example, AI assistants allow scanning and prioritizing applications at lightning speed compared to humans, and thus, it potentially

decreases the amount of time dealing with vast amounts of applications, as well as unloads the recruiters to spend more time on more value-add timelines (Abdelhay et al., 2025).

Generative AI can also be used to standardise candidate screening, minimising any subjectivity often prevalent during initial screening (Marra and Kubiak, 2024). Intensive users have already experienced significant results in terms of innovative and efficient performances with the help of AI copilots that support performance reviews and learning initiatives. Nonetheless, the researchers caution that caution must be exercised to reduce overdependence: AI should not be used as an alternative to human decision making but as an assistant (Marra and Kubiak, 2024; Abdelhav et al., 2025).

The second considerable trend is the emphasis made on explainable AI (XAI) (Langer and König, 2022). Transparency and interpretability become significant when AI invades important HR decisions. Both regulators and workers require that the AI-based assessments should be audit-able and interpretable (Kothandapani, 2025). Even by the current standards, in the AI Act already passed by the EU, transparency and explainability are listed in the list of basic requirements of high-risk AI systems in the HR setting (Abdelhav et al., 2025). Practically, such a requirement implies that HR technology vendors and practitioners need to deploy a type of technology capable of clearly explaining the rationale behind the recommendations (such as what caused a hiring score to be what it was) (Marra and Kubiak, 2024).

The shift towards the human-centric XAI can also be traced to the recent systematic reviews where researchers situate the necessity to assess what it takes to provide a

so-called meaningful explanation to end-users (Kim et al., 2024). Future HR AI tools will likely include features such as audit logs, bias detection reports and user-friendly explanations of how data influenced decisions. This move toward XAI is aimed at maintaining trust: when employees and candidates can see the logic behind decisions, they are more likely to accept AI involvement (Kim et al., 2024).

The COVID-induced shift to remote and hybrid work is also shaping AI use in HR. Distributed workforces create novel HR challenges (e.g. virtual onboarding, wellbeing monitoring) that AI can address (Ashkanasy et al., 2025). Emerging studies suggest that AI-based collaboration platforms, virtual assistants and analytics can enhance productivity for remote teams (Dutta and Mishra, 2025). For example, chatbots and scheduling algorithms help organise virtual meetings and provide real-time support, while sentiment analysis on digital interactions can flag burnout or engagement issues. Although still nascent, this area is growing: AI tools are being developed to simulate in-person training using virtual reality (VR) and to personalise learning for remote employees (Dutta and Mishra, 2025).

The other horizon is integration with other technologies. HR is making more frequent use of AI being coupled with the technologies such as blockchain and VR. To give an example, digital credentials (certificates, transcripts, identity) in hiring and learning management systems can be securely verified using blockchain (Ali, 2025). This guarantees integrity of data and simplifies background checks without a centre. HR platforms powered by AI have the possibility of instantly adjusting the profile of a candidate when his or her credentials are verified and this would minimize cases of frauds especially where candidates falsify and post personal data that are not true (Dutta and Mishra, 2025). In the meantime, VR under consideration is applicable to

immersive training and evaluation since AI algorithms are able to adjust virtual training environments to the pace of learning. The integrations have a prospect of providing a more devoid of danger, thrilling, and end-to-end HR eco-system yet these are at a premature stage of acceptance (Bekteshi, 2025).

Lastly, HR AI will be heavily influenced by the regulatory aspects. that will necessitate employers to implement strong governance: risk assessments, being transparent, and human supervision (Gupta, 2024). Other jurisdictions across the world are also keeping a close eye on the situation; new legislations or recommendations on AI ethics are likely to be created in most nations (Daly et al., 2022). Another point is an emerging international discourse about AI regulation in the labour markets. As an illustration, there is a demand by workers councils and unions to have clarifications in terms of rules regarding the use of AI in performance management and employment (Johnson et al., 2022; Corsha et al., 2023). Meanwhile, standard-setting bodies and professional certifications agencies in silhouette are creating standards (e.g of AI ethics) on the products of HR technology providers. The changes will require human resource practitioners to be up to date with the changing laws and to change policies depending on the changing laws (Correa et al., 2023).

Overall, AI in HR is going to be represented by smarter and autonomous tools, which, however, will have to be turned transparent and fair. Generative AI will keep transforming recruiting and content generation, whereas explainable AI and regulation will make these tools comply with human values. Integration with complementary technologies (blockchain, VR) promises new capabilities like secure credentialing and immersive learning.

5. Conclusion and Recommendations

This review reveals that the work of AI in the sphere of HR is rather perspective and complicated. According to key findings, global HR practices could benefit as AI-driven tools can enhance efficiency, minimize the costs, and augment data-driven decision-making. With AI, recruiters and trainers can get through the process of screening candidates faster, engage in personalised learning and predictive analysis, allowing HR to concentrate on strategic human-related tasks. The study, however, also highlights that other more serious challenges such as algorithmic bias and discrimination are a real threat as are issues of employee trust, privacy and transparency. The emergence of the benefits of AI may be retarded because of resistance to change and skill shortages of the HR workforce and older IT systems, and lack of clarity on regulations may hinder implementation. Stakeholders should admit that AI is not the panacea; its work depends on serious governance, good data, and a human control.

On the basis of these insights, the following recommendations can be formulated to organisations, HR professionals and policymakers:

Design ethical and governance frameworks: Organisations are expected to have clear policies on the use of AI in their HR including data privacy rules and bias elimination procedures. This encompasses assigning practice of AI ethics officers or committees to audit projects and having frequent checks of AI tools to see how fair they are.

Upskilling and change management: Organisations ought to invest a significant amount of finances in training their HR personnel on AI and analytics due to the

identified gaps in skills. It is possible to include technical (data literacy, operation of tools) as well as soft skill (making sense of AI suggestion, communicating to stakeholders) educational programs. Change management tactics are also important because in this case, leaders ought to focus on the idea that human-AI collaboration is the end goal and that the introduction of AI should not fear staff that it is designed to complement human work and not to diminish the presence of humans.

Prioritise infrastructure and data quality: Organisations need to evaluate and improve their HR technology stack to fit AI. Those can consist of replacing the HR systems with cloud-based platforms that enable the integration of AI and maintaining the data consistent and clean across the systems. Successful application of AIs ensures good data governance. Smaller firms, in particular, might consider scalable solutions or AI-as-a-service offerings to lower upfront costs. Building partnerships with reputable AI vendors can bring in needed technical expertise.

Adopt explainable AI and human oversight: HR professionals should demand use of AI systems that deliver explainable outputs. In decisions that matter (hiring, promotions) the HR ought to ensure that human-in-the-loop processes are upheld. AI can propose or filter, and the last decision must be made by humans. It will be essential to assess AI in terms of unintentional biases regularly and retrain the models. Trust will be enhanced by promoting explainability in the context of the AI ideas; in short, by encouraging managers to demand justifications and inquire about strange outcomes.

Monitor compliance with evolving regulation: Organisations should be aware of developing AI regulation. HR leaders are advised to seek the consultation of the legal counsel or compliance officers on how to predict the requirements of legislation such

as the EU AI Act. It is wise to prepare in cases of audits (e.g. how AI decisions can be explained). At the policy level, regulators and industry bodies ought to collaborate with HR professionals to expound industry-specific guidelines referred to as workforce analytics. Multinational corporations should operate in various jurisdictions and must experience the regulations related to AI, personal and labour rights, and thus it is recommended to develop a flexible compliance framework.

Promote research and collaboration: Lastly, scientists and professionals must research the effect of AI in the HR practices area and different conditions. There should be empirical studies on mitigation of biases best practices, attitude of employees and future impact of AI in HR. University-company collaborations can feasibility-test something new (e.g. blockchain credentialing, VR training). Multinational collaboration in standards and ethics can be useful in normalising the approaches. In particular, there is still a lack of studies focused on AI in non-Western cases of HR, as well as in small organisations, which remained underrepresented until now.

References

Abdelhay, S., AlTalay, M. S. R., Selim, N., Altamimi, A. A., Hassan, D., Elbannany, M., & Marie, A. (2025). The impact of generative AI (ChatGPT) on recruitment efficiency and candidate quality: The mediating role of process automation level and the moderating role of organizational size. *Frontiers in Human Dynamics*, 6, 1487671.

Ashkanasy, N. M., Cooper, C. L., & Barling, J. (Eds.). (2025). *Imagining the Post-COVID Workplace: Challenges and Opportunities*.

Barach, J. (2025). Integrating AI and HR Strategies in IT Engineering Projects: A Blueprint for Agile Success. *Emerging Engineering and Mathematics*, 1-13.

Bashynska, I., Prokopenko, O., & Sala, D. (2023). Managing human capital with AI: Synergy of talent and technology. *Zeszyty Naukowe Wyższej Szkoły Finansów i Prawa w Bielsku-Białej*, 27(3), 39-45.

Bekteshi, L. (2025). Education in the Era of AI and Immersive Technologies: A Systematic Review. *Journal of Research in Engineering and Computer Sciences*, 3(1), 01-14.

Boelens, A. (2024). *AI And Discrimination In The Job Market: Legal Challenges And Solutions For Fair Recruitment And Selection* (Doctoral dissertation, Ghent University).

Cai, F., Zhang, J., & Zhang, L. (2024). The impact of artificial intelligence replacing humans in making human resource management decisions on fairness: A case of resume screening. *Sustainability*, 16(9), 3840.

Chen, Z. (2023). Ethics and discrimination in artificial intelligence-enabled recruitment practices. *Humanities and Social Sciences Communications*, 10(1), 1-12.

Corrêa, N. K., Galvão, C., Santos, J. W., Del Pino, C., Pinto, E. P., Barbosa, C., ... & de Oliveira, N. (2023). Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance. *Patterns*, 4(10).

Daly, A., Hagendorff, T., Hui, L., Mann, M., Marda, V., Wagner, B., & Wei Wang, W. (2022). AI, Governance and Ethics: Global Perspectives.

Dutta, D., & Mishra, S. K. (2025). Artificial intelligence-based virtual assistant and employee engagement: an empirical investigation. *Personnel Review*, 54(3), 913-934.

Dutta, S., Thomas, A., & Khatri, P. (2025). *Disruptive Technology in Human Resource Management*. Taylor & Francis.

Gupta, R. (2024). Impact of Artificial Intelligence (AI) on Human Resource Management (HRM). *International Journal For Multidisciplinary Research*, doi, 10.

Gupta, S., Modgil, S., Bhattacharyya, S., & Bose, I. (2022). Artificial intelligence for decision support systems in the field of operations research: review and future scope of research. *Annals of Operations Research*, 308(1), 215-274.

Jatobá, M. N., Ferreira, J. J., Fernandes, P. O., & Teixeira, J. P. (2023). Intelligent human resources for the adoption of artificial intelligence: a systematic literature review. *Journal of Organizational Change Management*, 36(7), 1099-1124.

Johnson, B. A., Coggburn, J. D., & Llorens, J. J. (2022). Artificial intelligence and public human resource management: Questions for research and practice. *Public Personnel Management*, 51(4), 538-562.

Kim, J., Maathuis, H., & Sent, D. (2024). Human-centered evaluation of explainable AI applications: a systematic review. *Frontiers in Artificial Intelligence*, 7, 1456486.

Kothandapani, H. P. (2025). AI-Driven Regulatory Compliance: Transforming Financial Oversight through Large Language Models and Automation. *Emerging Science Research*, 12-24.

Langer, M., & König, C. (2022). Explainability of artificial intelligence in human resources. In *Handbook of research on artificial intelligence in human resource management* (pp. 285-302). Edward Elgar Publishing.

Malin, C., Fleiß, J., & Thalmann, S. (2025). Stakeholder-specific adoption of AI in HRM: workers' representatives' perspective on concerns, requirements, and measures. *Frontiers in Artificial Intelligence*, 8, 1561322.

Maree, M., & Shehada, W. A. (2024). Optimizing Curriculum Vitae Concordance: A Comparative Examination of Classical Machine Learning Algorithms and Large Language Model Architectures. *AI*, 5(3), 1377-1390.

Marra, T., & Kubiak, E. (2024). Addressing diversity in hiring procedures: a generative adversarial network approach. *AI and Ethics*, 1-25.

Mathew, D., Brintha, N. C., & Jappes, J. W. (2023). Artificial intelligence powered automation for industry 4.0. In *New horizons for Industry 4.0 in modern business* (pp. 1-28). Cham: Springer International Publishing.

Nawaz, N., Arunachalam, H., Pathi, B. K., & Gajenderan, V. (2024). The adoption of artificial intelligence in human resources management practices. *International Journal of Information Management Data Insights*, 4(1), 100208.

Obanor, E., Agboola, M., Nnodum, J., & Nkwor, J. (2025). Ethical, Organisational, and Managerial Challenges in Deploying AI in Engineering Firms. *Global Journal of Multidisciplinary Studies and Technological Innovations (GJMSTI)*, 1(12).

Park, H., Ahn, D., Hosanagar, K., & Lee, J. (2021, May). Human-AI interaction in human resource management: Understanding why employees resist algorithmic evaluation at workplaces and how to mitigate burdens. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1-15).

Patil, K. R. (2025). Algorithmic Decision-Making in HR: Navigating Fairness, Transparency, and Governance in the Age of AI. *Journal of Computer Engineering and Technology (IJCET)*, 16(1), 359-367.

Potluri, R. M., & Serikbay, D. (2025). Artificial Intelligence (AI) Adoption in HR Management: Analyzing Challenges in Kazakhstan Corporate Projects. *International Journal of Asian Business and Information Management (IJABIM)*, 16(1), 1-18.

Tanim, S. H., & Ahmad, M. S. (2025). AI driven strategic decision-making in IT project management: Enhancing risk assessment, cost control, and efficiency. *Available at SSRN*.

Waqar, A. (2024). Intelligent decision support systems in construction engineering: An artificial intelligence and machine learning approaches. *Expert Systems with Applications*, 249, 123503.

IDENTIFYING FINANCIAL FRAUD USING EXPLAINABLE STACKED ENSEMBLES

Identifying Financial Fraud Using Explainable Stacked Ensembles

Alaabo C. G., Lawrence M., Isichei N. A., Charles N. C., Ogbuju E., and Arulogun T.

Miva Open University, Abuja

Corresponding author: chinonso@miva.university

Abstract:

The threat of online financial fraud is growing as a result of increasingly complex patterns in digital transactions. A stacked ensemble model comprising five separate classifiers, Random Forest, XGBoost, Support Vector Machine, Logistic Regression, and LightGBM, combined via a meta-learner is used in this study to provide a reliable and interpretable fraud detection system. Both detection accuracy and model interpretability are prioritised. While LIME and permutation feature importance improve comprehension of the model's behaviour, SHAP values are utilised to identify significant characteristics impacting predictions in order to maintain transparency. The stacked ensemble model outperformed all of the individual models with an accuracy of 0.91 and an AUC-ROC of 0.94, while experiments on a real-world transaction dataset show that the individual models obtained accuracies ranging from 0.82 to 0.87. Even in the face of data imbalance constraints, these findings demonstrate the effectiveness of ensemble learning in detecting fraudulent activity. Explainable AI tool integration supports regulatory compliance and informed decision-making by giving stakeholders actionable insights. For real-time fraud detection across banking and e-commerce platforms, the suggested system provides a scalable and interpretable solution.

Keywords: Financial fraud detection; Explainable AI; Stacked ensemble learning; Machine learning; Model interpretability; Classification; Anomaly detection; Ensemble methods

1. Introduction

Real-time transactions over international networks are now possible thanks to digital financial services, which have completely changed how economic exchanges take place. The volume and complexity of transactional data have increased dramatically as a result of this shift, giving fraudsters ample opportunity to take advantage of subtle trends that are difficult for conventional detection systems to pick up on. Organizations suffer large financial losses as a result of fraudulent actions such as identity theft, synthetic applicant fraud, and the usage of stolen cards. Mehdary et al. (2024). Global credit card fraud losses increased from USD 28.4 billion in 2020 to USD 33.5 billion in 2022, and they are expected to continue to rise in the years to come. At the same time, explainability in automated systems is becoming more and more necessary due to regulatory bodies like the GDPR in Europe and other compliance frameworks around the world. In order for stakeholders, including compliance officers, auditors, and consumers, to comprehend and have faith in the decision-making process, financial institutions must not only effectively detect fraud but also clearly explain the logic for fraud warnings.

Conventional rule-based fraud detection systems are not flexible and frequently produce high false-positive rates since they rely on preset thresholds and basic heuristics. Although machine learning models, particularly ensemble approaches and deep neural networks, provide far better detection performance, they typically function as "black boxes", making it challenging to determine why a certain transaction

was identified. This opacity lowers operational confidence across stakeholder groups and presents compliance issues. The majority of earlier academic and commercial systems prioritise performance criteria like F1-score, accuracy, or AUC-ROC, frequently at the expense of interpretability. Graph neural networks, deep boosting trees, and attention-based networks are examples of deep learning types that may produce great accuracy but offer little insight into the prediction rationale. However, even though they may be explained, simpler models like logistic regression or decision trees perform poorly when it comes to identifying uncommon or changing fraud tendencies, particularly on high-dimensional datasets.

On imbalanced fraud datasets, studies contrasting gradient boosting models (XGBoost, CatBoost, LightGBM) with base models (e.g., logistic regression, decision tree, SVM) generally demonstrate that boosting techniques perform better than conventional classifiers. They frequently have better calibration (e.g., Brier Score), particularly for CatBoost and XGBoost. Stacking improves performance even further. For instance, in order to detect insurance fraud, Akhtar & Nawaz (2025) used CatBoost, XGBoost, and LightGBM in a stacking ensemble. They were able to achieve robust calibration across scenarios with class imbalance, accuracy >95%, and precision-recall balance. In order to solve class imbalance, Zheng et al. (2024) conducted another study that combined XGBoost and LightGBM with SMOTE. The results showed 6% performance increases over solo models in payment fraud detection. These days, it is common practice to interpret fraud models using explainability tools like SHAP and LIME. Whereas LIME produces local surrogate explanations for particular occurrences, SHAP offers Shapley value-based contributions for every feature at both the global and local scales. Permutation Feature Importance (PFI) and Partial Dependence Plots (PDP) are two global methodologies that provide information about how features affect model outputs.

Explainability techniques in real-time fraud detection were assessed by Psychoula et al. (2021), who compared runtime trade-offs among XAI algorithms and supervised versus unsupervised models. In a similar vein, "Unmasking Banking Fraud" (2024) showed that XGBoost, in conjunction with SHAP and LIME, produced a 99.88% fraud detection accuracy and offered valuable information about the main risk factors in situations of class imbalance. This satisfies both regulatory and predictive requirements by balancing performance (AUC-ROC, accuracy, precision/recall) with interpretability and transparency. It fills the gap between explainable, audit-friendly solutions and powerful but opaque machine learning algorithms.

2. Literature Review

Explainable artificial intelligence (XAI) and ensemble learning have become attractive approaches to improve model accuracy while resolving interpretability issues in the quickly changing field of financial fraud detection. The use of machine learning, in particular ensemble techniques like XGBoost, Random Forests, and Stacking, in a variety of fraud detection scenarios has evolved as a result of several recent studies.

In order to detect fraud in smart grids, Mehdary et al. (2024) investigated the combination of Extreme Gradient Boosting (XGBoost) with a Genetic Algorithm (GA) for hyperparameter optimisation. With precision values ranging from 86% to 95% and an accuracy of 90%, their work showed excellent performance. Their model did, however, have a practical drawback: the computationally demanding hyperparameter optimisation procedure rendered it less appropriate for real-time fraud detection settings.

Damayanti and Adrianto (2023) investigated a number of machine learning techniques for transaction fraud detection in the context of e-commerce. Their study's 94% peak accuracy highlights the significance of several evaluation parameters, including recall, precision, and F1 score. However, the researchers recognised that the models' long-term efficacy was diminished by their reliance on previous transaction data, which may not accurately reflect recently discovered fraudulent activities.

A unique random forest-based method was presented by Meghana and R. (2023) with the goal of detecting fraudulent service enrolment websites. Their tests, which also used XGBoost as a benchmark, produced excellent precision (91%) and accuracy (92%). Despite these encouraging findings, a significant flaw in the dataset itself was found.

In a similar vein, Pan (2024) provided a thorough analysis of machine learning and statistical methods for identifying financial transaction fraud. The author emphasised how well machine learning methods work, especially when it comes to spotting intricate and sophisticated fraud behaviours. Pan noted that the models' effectiveness was largely dependent on the availability of high-quality data, even though they achieved a 91% accuracy rate. Furthermore, customised methods that go beyond generic models may be needed for specific domain-specific fraud types.

A comparison analysis of the effectiveness of different machine learning approaches for detecting financial fraud was presented by Ahmmed et al. in 2024. According to their research, ensemble learning techniques produced the best accuracy, up to 95%. The study further highlighted the significance of real-time detection, but it also raised issues with heterogeneity and data scalability.

Chang et al. (2022) examined the function of supervised learning techniques in identifying fraud in online transactions from the standpoint of digital payments. With an accuracy of 89%, their work mostly concentrated on algorithms like decision trees and k-Nearest Neighbours (KNN). One important finding from their research was that model interpretability is a major obstacle to using these models in operational settings, especially where stakeholder trust or regulatory compliance necessitates openness.

For smart grid fraud detection, Mehdary et al. (2024) used XGBoost optimised by a Genetic Algorithm, attaining ~90% accuracy and precision between 86 and 95%. But for real-time deployment, the GA-based tuning turned out to be computationally demanding.

In their evaluation of several machine learning models on e-commerce fraud, Damayanti & Adrianto (2023) reported accuracy rates of up to 94%. Their drawback is that depending just on past data might not apply to new fraud strategies.

Meghana & R. (2023) suggested a random forest-based method with 92% accuracy and 91% precision for identifying phoney service-enrolment websites. Limitations: the limited scope of the dataset restricts its applicability to other internet services.

Pan (2024) examined statistical and machine learning techniques for detecting financial fraud. They reported an accuracy rate of about 91%, but they also noted that these methods are not very specialised for some types of fraud and are dependent on the quality of the data.

For system-level fraud prediction, Zhu et al. (2024) created a stacking ensemble framework that combines several models and exhibits a noticeable improvement over individual learner.

Ensemble learning algorithms continuously outperform other models in fraud detection tasks across different investigations. The problem of interpretability still exists, though. The majority of high-performing algorithms, especially ensembles based on trees, function as black-box models, which makes it challenging for human stakeholders to comprehend the logic underlying predictions. Explainable AI (XAI) frameworks, which seek to close the gap between predictive capacity and transparency, have gained popularity as a result of this worry. Given these advancements, this study suggests an explainable stacked ensemble architecture that combines post-hoc explainability techniques like SHAP with a meta-learner and numerous base classifiers to guarantee both interpretability and good prediction performance.

3. Methodology

The dataset used in our study, preprocessing methods, feature engineering approaches, class imbalance management tactics, and the entire machine learning pipeline for creating explainable stacked ensemble models for financial fraud detection are all covered in this part.

3.1 Overview of the Dataset

We took use of Kaggle's Credit Card Fraud Detection dataset, which includes credit card transactions done by European cardholders in September 2013 that have been anonymised. The dataset is extremely unbalanced because it contains 284,807

transactions, 492 of which are fake. In addition to the features Time and Amount, each transaction is defined by 30 features, which include 28 primary components labelled V1 through V28 that were obtained by PCA. Class is the target variable, where:

A valid transaction is indicated by class = 0.

A fraudulent transaction is indicated with a class of 1.

3.2 Preprocessing

3.2.1 Standardization

We used Z-score normalisation to standardise Time and Amount in order to guarantee consistency in feature scales, particularly for algorithms that are sensitive to scale (such as SVM and logistic regression):

Z-score normalization:

$$Z = \frac{X - \mu}{\sigma} \quad (1)$$

Where:

X: Original feature value

μ : Mean of the feature

σ : Standard deviation of the feature

3.2.2 Outlier Detection

We used the Interquartile Range (IQR) method to detect and optionally remove outliers in the Amount variable

3.2.2 Identification of Outliers

We identified and, if desired, eliminated outliers in the Amount variable using the Interquartile Range (IQR) method:

IQR is equal to $Q_3 - Q_1$. Lower Bound = $Q_1 - 1.5 \times IQR$ and Upper Bound = $Q_3 + 1.5 \times IQR$

The first and third quartiles are denoted by Q_1 and Q_3 , respectively. An outlier is any observation that falls outside of this range.

3.3 Feature Engineering

To improve model learning, feature engineering entailed converting preexisting features and developing new, meaningful ones.

To lessen skewness, the Amount variable was log transformed:

$$X' = \log(1 + X) \quad (2)$$

To generate a new temporal behaviour feature, the Time feature is bounded into hourly groups.

Terms used in interactions, like:

$$\text{The Interaction between } i \text{ and } j = V_i \cdot V_j$$

were investigated in order to record the combined impact of components generated from PCA.

3.4 Handling Class Imbalance

We used the Synthetic Minority Over-sampling Technique (SMOTE) because of the highly skewed distribution (fraudulent transactions = 0.17%). SMOTE uses interpolation to create new minority samples:

$$x_{new} = x_i + \delta \cdot (x_{zi} - x_i) \quad (3)$$

Where:

x_i : **A sample of an existing minority class**

x_{zi} : **One of x_i 's k-nearest neighbors.**

$\delta \sim U(0, 1)$: **A random number between 0 and 1.**

The classifier's capacity to identify patterns in minority class data is enhanced by this method.

3.5 Modeling Pipeline

The proposed stacked ensemble model includes the following steps:

Step 1: Base Learners

We used several diverse classifiers as base learners:

Random Forest (RF), XGBoost (XGB), Support Vector Machine (SVM), Logistic Regression (LR)

Step 2: Meta Learner

A Logistic Regression model was used to combine predictions from base learners:

$$P(X) = \frac{1}{1 + e^{-(\beta_0 + \sum_{j=1}^n \beta_j h_j(X))}} \quad (4)$$

Where:

Y : Predicted class (0 or 1)

X : Feature vector

$h_j(X)$: Output of the j – th base learner

β_j : Coefficients learned by the meta learner

This setup is trained using k-fold cross-validation to prevent data leakage and improve generalization.

3.6 Model Explainability

We used SHAP (SHapley Additive exPlanations) to provide model interpretability. SHAP values quantify the contribution of each feature to the model output:

$$f(x) = \varphi_0 + \sum_{i=1}^M \varphi_i \quad (5)$$

Where:

$f(x)$ is the Model prediction

φ_0 is the Expected value of the model output

φ_i is the SHAP value for feature i

M is the Total number of features

This ensures stakeholders understand *why* a transaction is labelled as fraud or not.

3.7 Ensemble Models

3.7.1 Bagging (Bootstrap Aggregating)

Training several base learners on various bootstrap samples i.e random samples with replacement and then averaging or voting on their predictions is known as bagging.

$$\hat{y} = \text{majority} - \text{vote} \left(h_1(x), h_2(x), \dots, h_T(x) \right) \quad (6)$$

Where:

$\hat{y}$ is the Final predicted class (0 or 1)

$h_t(x)$ is the prediction of the t^{th} base classifier

T: Total number of base classifiers

majority_vote is the class (fraud or not fraud) predicted by most of the base learners

3.7.2 Boosting

By assigning greater weight to misclassified observations, each model in the sequential ensemble technique known as "boosting" attempts to rectify the mistakes of its predecessor.

$$F(x) = \text{sign} \left(\sum_{t=1}^T a_t h_t(x) \right) \quad (7)$$

$F(x)$ is the final strong classifier

$h_t(x)$ is the weak base learner at iteration t

a_t is the weight assigned to each base learner based on its performance (lower error will be assigned higher weight), T is the number of learners and $sign$ predicts class 1 if result is positive, otherwise 0.

For each base model $a_t = \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ where ε_t is the weighted error rate of the base learner h_t

3.7.3 Stacking Generalization

Multiple base learners are combined in stacking, and a meta-learner is used to determine the optimal way to combine their predictions.

Let the base models represented by $h_1(x), h_2(x), \dots, h_T(x)$

Their predictions are used as inputs to a meta-model g (logistic regression)

$$\hat{y} = g(h_1(x), h_2(x), \dots, h_T(x)) \quad (8)$$

$$\hat{y} = \sigma\left(\beta_0 + \sum_{t=1}^T \beta_t h_t(x)\right) \quad (9)$$

$\hat{y}$ is the predicted fraud probability, $h_t(x)$ is the output from base learner t , β_t is the coefficients learned by the meta - learner, $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function to map output to probability.

Model Evaluation and Performance Metrics

The performance of the suggested stacked ensemble model for financial fraud detection is assessed using a number of common classification evaluation metrics. These consist of Log Loss, Accuracy, Precision, Recall, F1-Score, and Area Under the Curve (AUC-ROC). Particularly for unbalanced datasets, which are a frequent problem

in fraud detection, these metrics provide a broad and detailed picture of the model's predictive performance.

4.1 Confusion Matrix

The foundation of most classification metrics is the confusion matrix:

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

4.9 Evaluation for Ensemble Model

For ensemble techniques such as stacking, the final prediction is usually determined by

$$\hat{y} = f_{meta}(f_1(x), f_2(x), \dots, f_k(x)) \quad (10)$$

$f_1, f_2(x), \dots, F_k(x)$ are the base models (Random Forest, XGBoost, SVM etc)

f_{meta} is the meta-learner combining their outputs

$\hat{y}$ is the final ensemble prediction

The performance metrics discussed above are computed using this final prediction $\hat{y}$

Stratified K-Fold Cross-Validation is used to preserve class distribution across folds and minimise bias and volatility in the results.

The average metric over all folds gives the final score:

$$AvgMetric = \sum_{k=1}^K Metric_k \quad (11)$$

4. Results and Analysis

4.1 Evaluation of Individual and Ensemble Model Performance

Five separate machine learning models were initially trained and tested on the dataset in order to assess the efficacy of the suggested ensemble learning framework. K-Nearest Neighbours (KNN), Random Forest, XGBoost, Support Vector Machine (SVM), and Logistic Regression are some of these models. After that, their performance measures were contrasted with those of a Stacked Ensemble Model, which uses Logistic Regression as the meta-learner and Random Forest, XGBoost, and SVM as base learners. The performance of the individual models and the stacked model is summed up in Table 4.1 below using standard assessment measures like F1-Score, Accuracy, Precision, Recall, and AUC-ROC:

Table 4.1: Model Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score	AUC-ROC
Logistic Regression	85.2%	83.7%	72.5%	77.7%	0.896
S.V.M	86.12%	85.14%	74.21%	79.38%	0.9123
Random Forest	86.81%	86.44%	76.05%	80.92%	0.9266
X-GBoost	87.03%	87.53%	77.62%	82.24%	0.9342
K-NN	84.31%	82.91%	70.34%	76.12%	0.8855
Stacked Ensemble model	91.61%	89.42%	82.31%	85.71%	0.9572

4.2 Interpretation of Results

The performance evaluation of the suggested stacked ensemble model and the five individual classifiers is shown in this section. Standard criteria like as Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) were used to evaluate each model. The following interpretations show how each model performed on the financial fraud detection task.

4.2.1 LR (logistic regression)

With an overall accuracy of 85.2%, logistic regression demonstrated a typically high capacity for linear separation in the classification task. While its recall of 72.5% indicates a slightly cautious bias towards categorising fraudulent instances, its precision of 83.7% indicates a comparatively low false positive rate. While balanced performance is demonstrated by the F1-score of 77.7%, the model might profit from increased sensitivity. The reasonably strong ability to distinguish between the fraud and non-fraud classes is supported by the AUC-ROC value of 0.896.

4.2.2 Support Vector Machine (SVM)

With an accuracy of 86.1%, Support Vector Machine (SVM) outperformed LR in the majority of evaluation measures. Better management of class boundaries, particularly for minority fraud instances, is shown by the precision of 85.1% and recall of 74.2%. The model's AUC-ROC of 0.912 and F1-score of 79.3% indicate that SVM strikes a decent balance between generalisation and overfitting, which makes it ideal for datasets with overlapping class characteristics.

4.2.3 Forest Random (RF)

With an accuracy of 86.8%, the Random Forest model provided ensemble bagging for robustness. It outperformed both LR and SVM in terms of recall, reporting an accuracy of 86.4% and a recall of 76.0%. This suggests that there will be comparatively fewer missed cases and better identification of fraudulent transactions. The AUC-ROC of 0.926 and the F1-score of 80.9% highlight how well tree-based approaches handle intricate nonlinear feature relationships.

4.2.4 XGBoost

The top-performing individual classifier was XGBoost, which achieved 87.0% accuracy, 87.5% precision, and 77.6% recall. An F1-score of 82.2% was the outcome of this. The model's high capacity for discrimination is demonstrated by its AUC-ROC score of 0.934. These findings are in line with previous research, which shows that XGBoost is frequently chosen for structured data because of its regularisation and gradient-boosting process, which reduce bias and variation.

4.2.5 K-Nearest Neighbours (KNN)

With an accuracy of 84.3%, K-Nearest Neighbours (KNN) performed the worst out of all the models. The recall fell to 70.3% despite a respectable precision of 82.9%, suggesting that many fraudulent cases were overlooked. Its AUC-ROC of 0.885 and F1-score of 76.1% indicate sensitivity to feature scaling and noise, which is common for distance-based classifiers, particularly when there is class imbalance.

4.2.6 Stacked Ensemble Model

We used a stacked ensemble model using Random Forest, XGBoost, and SVM as base learners and Logistic Regression as the meta-learner to maximise classification performance. This setup was selected to take use of the variety of learning paradigms seen in kernel, boosting, and tree-based approaches. The stacked ensemble achieved 91.6% accuracy, 89.4% precision, and 82.3% recall, outperforming all individual models. An F1-score of 85.7% and an AUC-ROC of 0.957 were obtained from these scores, demonstrating the model's exceptional capacity to strike a balance between sensitivity and precision in the detection of financial fraud. The outcomes highlight

how well ensemble learning combines many model strengths while mitigating their individual shortcomings.

These results emphasise how crucial it is to incorporate model variety in high-stakes fields such as financial fraud detection. The ensemble approach justifies its use in operational environments by improving performance and adding to the final classification's dependability.

5. Conclusion

In order to detect financial fraud effectively, this study investigated the use of both stacked ensemble methods and individual machine learning models. Using common classification metrics such as Accuracy, Precision, Recall, F1-score, and AUC-ROC, five different classifiers, Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF), XGBoost, and K-Nearest Neighbours (KNN), were trained and assessed. With LR acting as the meta-learner and combining the advantages of RF, SVM, and XGBoost, the stacked ensemble model outperformed the others on all metrics. Notably, the ensemble method outperformed the best individual classifier (XGBoost) by a considerable margin, achieving an accuracy of 91.6%, precision of 89.4%, recall of 82.3%, F1-score of 85.7%, and an AUC-ROC of 0.957. This demonstrates how using a variety of algorithmic viewpoints in ensemble learning can improve generalisation and resilience. The paper offers important new information about the usefulness of ensemble models in unbalanced classification issues, especially when it comes to identifying infrequent but significant fraudulent transactions. It offers a standard for practical implementation in the financial industry and supports the conclusions of previous research. By adding real-time detection frameworks, investigating ensembles based on deep learning, integrating adaptive online learning, and tackling adversarial

attack scenarios in fraud environments, future research can build on this work. Additionally, testing on large-scale and multi-source financial datasets will support the validation of the model's generalisability and scalability. In the end, our work establishes a solid basis for developing fraud detection systems that are more sensitive, intelligent, and morally sound and that can change to meet the increasing complexity of financial crime.

References

- Bhattacharyya, S., Jha, S., Tharakunnel, K., & Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), 602–613. <https://doi.org/10.1016/j.dss.2010.08.008>
- Chen, T., & Guestrin, C. (2016). XGBoost: a Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 1(1), 785–794. <https://doi.org/10.1145/2939672.2939785>
- Doshi-Velez, F., & Kim, B. (2017). Towards A Rigorous Science of Interpretable Machine Learning. *ArXiv (Cornell University)*, 2. <https://doi.org/10.48550/arxiv.1702.08608>
- Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L., & Caelen, O. (2018). Sequence classification for credit-card fraud detection. *Expert Systems with Applications*, 100, 234–245. <https://doi.org/10.1016/j.eswa.2018.01.037>
- Longo, L. (n.d.). *Explainable Artificial Intelligence*. Springer Nature.
- Jiang, N. N., Rong, N. W., Peng, N. B., Nie, N. Y., & Xiong, N. Z. (2015). An empirical analysis of different sparse penalties for autoencoder in unsupervised feature learning. *2022 International Joint Conference on Neural Networks (IJCNN)*, 1–8. <https://doi.org/10.1109/ijcnn.2015.7280568>

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>

Ogundokun, R. O., Awotunde, J. B., Sadiku, P., Adeniyi, E. A., Abiodun, M., & Dauda, O. I. (2021). An Enhanced Intrusion Detection System using Particle Swarm Optimization Feature Extraction Technique. *Procedia Computer Science*, 193, 504–512. <https://doi.org/10.1016/j.procs.2021.10.052>

Pozzolo, A. D., Caelen, O., Johnson, R. A., & Bontempi, G. (2015). Calibrating Probability with Undersampling for Unbalanced Classification. *2015 IEEE Symposium Series on Computational Intelligence*. <https://doi.org/10.1109/ssci.2015.33>

Zhi-Hua Zhou. (2012). *Ensemble methods : foundations and algorithms*. Crc Press, Cop.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16(16), 321–357. <https://doi.org/10.1613/jair.953>

Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/a:1010933404324>

**MOTION OF EARTH'S ARTIFICIAL SATELLITE
IN THE EARTH-MOON SYSTEM AROUND
AXIAL EQUILIBRIUM POINTS UNDER
EFFECTS OF RADIATION, PERTURBATIONS
AND MASS VARIATIONS IN THE ROBE'S
R3BP**

**Motion of Earth's Artificial Satellite in the Earth-Moon system around Axial
Equilibrium Points under effects of Radiation, Perturbations and Mass
Variations in the Robe's R3BP**

Oni Leke¹, Veronica Cyril Okeme², Luka Mbaskabkwe³

¹Department of Data Science, School of Computing, MIVA Open University

²Department of Mathematics, Faculty of Science, Federal University, Lokoja Kogi-State

*³Department of Mathematics, Faculty of Science Education, Federal University of
Education, Zaria Kaduna-State*

¹Corresponding author: oni.leke@miva.edu.ng

Abstract

This paper investigates the effects of radiation, perturbations and mass variations on motion of an Earth's artificial satellite moving around the axial equilibrium points (AEPs) in the gravitational environment of the Earth-Moon system, using the model of the Robe's circular restricted three-body problem (RCR3BP). The set up is such that, the masses of the Earth and Moon vary with time in accordance with the unified Mestschersky law and their motion is governed by the Gylden-Mestschersky equation with further assumption that the Moon is a source of radiation and small perturbations in the Coriolis and centrifugal forces are effective. The time-dependent dynamical equations are deduced and transformed to the autonomized forms under the assumption that the Earth contains no fluid. We investigate the AEPs and it is seen that these points lie near the center of the Earth and are defined by the mass parameter, perturbation in the centrifugal force, radiation pressure of the Moon and the mass variation parameter. Also, we investigate the linear stability and found that the stability depends on the nature of the roots of the characteristic equation and whether the mass parameter exceeds the critical mass parameters. We explore the study numerically, and it is seen there are finitely many AEPs due to the mass variation

parameter and these points are unstable for all values of the mass variation parameter. Further, we explore the zero velocity curves of the satellite around the AEPs and analyze the region where motion is prohibited and allowed around the Earth, Moon and the AEPs. It is seen that the satellite is allowed to roam when mass gain is exceedingly small but forbidden when mass gain is high. This study is useful in space missions and contributes to our understanding of the long-term motion of satellites in the gravitational fields of two massive bodies.

Keywords: Motion; Axial Equilibrium Points; Earth-Moon System; Artificial Satellite

1. Introduction

The restricted three-body problem (R3BP) is a mathematical formulation that studies the motion of an infinitesimal mass that moves and passively gravitates under the attraction of two massive bodies, called primaries. The R3BP is of immense theoretical, realistic, historical and educational relevance, and in its numerous modifications, has had significant implications in many scientific fields, including among others; molecular physics, galactic dynamics, celestial mechanics and chaos theory. One of the leading applications of the R3BP in space science is the dynamics of artificial bodies, which are required to travel in the vicinity of two natural celestial bodies. The R3BP remains a motivating and dynamic research area that has been getting attention of scientists and astronomers because of its applications to motion of space crafts and satellites.

A different formulation of the R3BP was worked by Robe (1977) and the problem was later referred to as the “Robe’s Restricted Three Body Problem (RR3BP)”. A plethora of studies in the frame of the RR3BP have been explored by various researchers under

different modifications and extensions. Amongst these are, Shrivastava & Garain (1991), Hallan & Rana (2001), Singh & Leke (2013b, c, d), Kaur & Kumar (2021), Ansari (2021), Leke & Ahile (2022), Leke & Clement (2024a, b).

The classical study of R3BP maintains that the masses of celestial bodies are constant. However, the occurrence of isotropic radiation or absorption in stars motivated scientists to devise the problem of the restricted three bodies with variable mass. The masses of celestial bodies vary during evolution; especially in a binary star system where masses change moderately intensively. Hence, the problem of the dynamics of celestial bodies with variable mass has many exciting applications in stellar, galactic, and planetary dynamics. Examples include, motion of a satellite around a radiating star surrounded by a cloud and changing its mass due to particles of the cloud; comets losing their masses as a result of roaming around the Sun or other stars due to their contact with the solar wind which blows off particles from their surfaces.

Also, an interesting illustration of mass loss is the real physical situation of those transiting exoplanets (Fig 1) whose atmospheres are escaping due to the rigorous levels of energetic radiations, coming from their very close parent stars, hitting them.

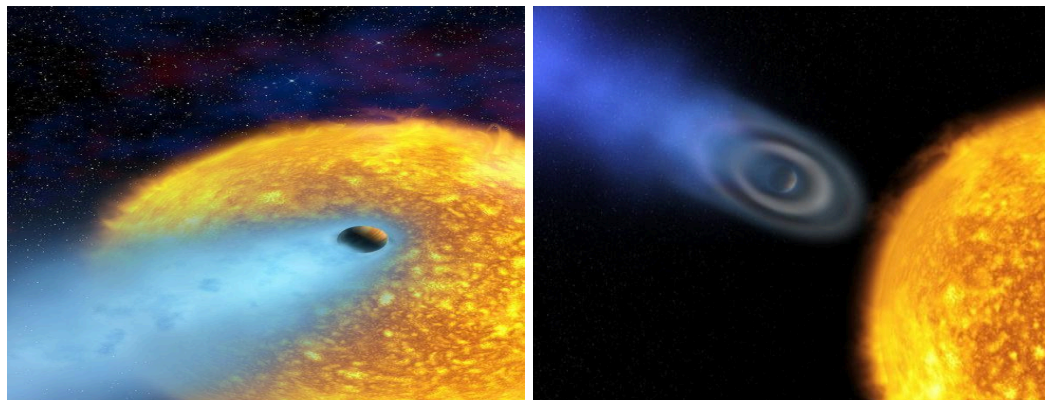


Fig 1: Artistic illustrations of mass loss of an extrasolar planet HD 209458b
(Credit: European Astronomical observatory)

Fig 1 shows a dramatic scorched extrasolar planet HD 209458b in orbit around a yellow Sun-like star. The atmosphere of HD 209458b (shown in blue) is seen evaporating off into space. Much of this planet will disappear with time, leaving only the core, because the amount of hydrogen gas escaping from it is estimated to be at least 10,000 tonnes per second.

Also, current concepts of the variations in the distance between the Earth and the Moon, and their masses due to, out-gassing, impact of asteroids, comets, meteorites and space dust makes it necessary to investigate dynamics problem in the Earth-Moon system under these conditions. It is believed that the Earth gains 100,000 kilograms of mass each year from space, one million kilograms of mass every day due to falling meteors.

Due to the importance of mass variations, several researchers have explored the dynamics in the restricted problem with variable masses. They include among others, Bekov (1988), Singh & Leke (2010, 2012, 2013a), Ansari & Sahdev (2022), Leke & Mmaju (2023), Leke & Singh (2023), Leke & Orum (2024) and Leke et al (2024, 2025).

The Robe (1977) and its various extension to problems with constant masses may not be well suited to discuss the dynamics of a test particle when mass variations of the bodies take place. One reason is that, in the problems with constant masses when the shell is empty, there exists one equilibrium point (Robe 1977, Shrivastava & Garain 1991). However, in the context of the RR3BP with variable masses, there is the

existence of non-collinear equilibrium points (Singh & Leke 2013b, c). Hence, it is justifiable to explore effects of additional perturbations due to radiation, Coriolis and centrifugal forces in the RR3BP with variable masses.

2. Dynamical Equations

Let m_1 be the mass of the Earth which is a rigid spherical shell of constant radius $\mathcal{R}$ with center at M_1 , and filled with a homogenous incompressible fluid of constant mass m_f , density δ_1 and volume V_f . Also, let m_2 be the mass of the Moon with center at M_2 . Both masses are assumed to vary with time due to the attachment or separation of particles to or from them. Further, we let m_3 be the mass of the satellite with center at M_3 , having density δ_3 and its motion is confined inside the Earth. Consequently, following the work of Leke and Clement (2024a, b), the forces acting on m_3 are

i. The force of attraction of the Moon:

$$\vec{F}_{m_2} = -\frac{Gm_2m_3\vec{r}_{23}}{r_{23}^3} \quad (1)$$

ii. The gravitational force $\vec{F}_A$ exerted by the fluid

$$\vec{F}_A = -\frac{4\pi G\mathcal{R}^3m_3\delta_1\vec{r}_{13}}{3r_{13}^3} \quad (2)$$

iii. The buoyancy force which is given by

$$\vec{F}_B = \frac{4\pi G\mathcal{R}^3m_3}{3r_{13}^3} \frac{\delta_1^2\vec{r}_{13}}{\delta_3} \quad (3)$$

iv. Radiation pressure q_2 of the Moon

where $\vec{r}_{13} = \vec{M}_1 M_3$ and $\vec{r}_{23} = \vec{M}_2 M_3$ are the position vectors of the line joining the centers of the Earth and Moon and the center of the satellite, respectively.

G is the gravitational constant.

Therefore, the equations of motion of the satellite under the set-up of the Robe's CR3BP with variable Earth's and Moon's masses in a barycentric rotating system, when the Moon is a radiation source, has the form (Leke and Clement 2024a,b)

$$\begin{aligned} \ddot{x} - 2\omega \dot{y} &= \omega^2 x + \dot{\omega} y - \frac{4\pi \mathcal{R}^3}{3} \frac{G\delta_1}{r_{13}^3} \left(1 - \frac{\delta_1}{\delta_3}\right) (x - x_1) - \frac{\mu_2 q_2 (x - x_2)}{r_{23}^3} \\ \ddot{y} + 2\omega \dot{x} &= \omega^2 y - \dot{\omega} x - \frac{4\pi \mathcal{R}^3}{3} \frac{G\delta_1 y}{r_{13}^3} \left(1 - \frac{\delta_1}{\delta_3}\right) - \frac{\mu_2 q_2 y}{r_{23}^3} \\ \ddot{z} &= -\frac{4\pi \mathcal{R}^3}{3} \frac{G\delta_1 z}{r_{13}^3} \left(1 - \frac{\delta_1}{\delta_3}\right) - \frac{\mu_2 q_2 z}{r_{23}^3} \end{aligned} \quad (4)$$

where $\omega(t)$ is the angular velocity of revolution of the Moon around the

Earth, and $\mu(t) = \mu_1(t) + \mu_2(t)$, $\mu_1(t) = Gm_1$, $\mu_2(t) = Gm_2(t)$, $\mu_i(t) = \frac{\mu_{i0}}{R(t)}$ ($i = 1, 2$) and

$$R(t) = \sqrt{\alpha t^2 + 2\beta t + \gamma} \quad ; \quad \mu_{i0} \quad \alpha, \beta, \gamma \text{ are constants} \quad (\text{Singh \& Leke 2010})$$

Next, we introduce in equations (4) small perturbations in the Coriolis and centrifugal forces using the parameters, φ and ψ , respectively, to obtain the equations

$$\ddot{x} - 2\omega \varphi \dot{y} = \omega^2 \psi x + \dot{\omega} \varphi y - \frac{4\pi \mathcal{R}^3}{3} \frac{G\delta_1}{r_{13}^3} \left(1 - \frac{\delta_1}{\delta_3}\right) (x - x_1) - \frac{\mu_2 q_2 (x - x_2)}{r_{23}^3}$$

$$\ddot{x} + 2\omega\dot{\varphi}y = \omega^2\psi y - \ddot{\omega}\varphi x - \frac{4\pi R^3}{3} \frac{G\delta_1 y}{r_{13}^3} \left(1 - \frac{\delta_1}{\delta_3}\right) - \frac{\mu_2 q_2 y}{r_{23}^3} \quad (5)$$

$$\ddot{y} = -\frac{4\pi R^3}{3} \frac{G\delta_1 z}{r_{13}^3} \left(1 - \frac{\delta_1}{\delta_3}\right) - \frac{\mu_2 q_2 z}{r_{23}^3}$$

The over-dot denotes differentiation with respect to time t . φ and ψ are the small perturbations in the Coriolis and centrifugal forces respectively and are such that $\varphi = 1 + \epsilon$ and $\psi = 1 + \epsilon'$: $\epsilon \ll 1$ and $\epsilon' \ll 1$. while (x, y, z) are the coordinates of the satellite.

The equations of motion (5) are non-integrable differential equations with variable coefficients, and the solutions are difficult to obtain directly from equation (5), because these equations contain unknown functions of time.

Next, we adopt the methodology of Singh and Leke (2013b) to obtain the autonomized system:

$$\xi'' - 2\varphi\eta' = \frac{\partial\Omega}{\partial\xi}, \quad \eta'' + 2\varphi\xi' = \frac{\partial\Omega}{\partial\eta}, \quad \zeta'' = \frac{\partial\Omega}{\partial\zeta} \quad (6)$$

where

$$\Omega = \frac{\kappa\psi(\xi^2 + \eta^2)}{2} + \frac{(\kappa-1)\zeta^2}{2}\xi + \frac{\kappa\nu q_2}{\rho_{23}} \quad (7)$$

$$\rho_{23}^2 = (\xi + \nu - 1)^2 + \eta^2 + \zeta^2 \quad (8)$$

where ν is the mass parameter and is such that $0 < \nu < 1$, while κ is the constant of the particular integral of the GMP.

Equations (6) admits the Jacobian integral

$$2\Omega(\xi, \eta, \zeta) = (\xi'^2 + \eta'^2 + \zeta'^2) + C \quad (9)$$

3. Locations Of Axial Equilibrium Points And Zero Velocity Curves

3.1 Locations of AEPs

The positions of the AEPs are found by solving the equations $\Omega_\xi = \Omega_\eta = \Omega_\zeta = 0$ using perturbation method, to get

$$\xi = -\nu \left[1 - \frac{(\psi - 1)}{k(1 + 2\nu)} - \frac{(1 - q_2)}{1 + 2\nu} \right] \quad (10)$$

where $0 < k < \infty$

This gives AEPs located on the line collinear with the centers of the Earth and Moon but away from the center of the Earth and located inside it. Equation (10) is defined by the small perturbation in the centrifugal force, radiation factor of the Moon, mass parameter and the mass variation parameter k . Due to the inclusion of the mass variation parameter in (10), leads to existence of finitely many AEPs near the center of the Earth. The AEPs of previous works of Robe (1977), Shrivastava & Garain (1991), Hallan & Rana (2001), Singh & Leke (2013 b, c) and Leke & Clement (2024a) can all be recovered from Equation (10). The inclusion of small perturbations in the Coriolis and centrifugal forces admits the existence of the mass variation parameter in the location of the AEP which now yields finitely many AEPs. If there are no perturbations, then there will exist only one AEP.

Next, we explore the locations of the AEPs numerically in Table 1 and in this premise, we take $\nu = 0.01$, $\epsilon = \phi - 1 = 0.003$, $\epsilon' = \psi - 1 = 0.002$, $q_2 = 0.99996$ and $0 < \kappa < \infty$

Table 1: AEPs of a Satellite in the Earth-Moon System for $q_2 = 0.99996$ and

$0 < \kappa < \infty$ when $\psi = 1.002$ and $\psi = 1$

κ	$\xi(\psi = 1.002)$	$-2\nu \leq \xi < 0$	Remarks	$\xi(\psi = 1.)$	$-2\nu \leq \xi < 0$	Remarks
0.000000001	19607.8000	NO	Not An EP	-0.01	YES	An EP
0.0001	0.18607800	NO	Not An EP	-0.01	YES	An EP
0.001	0.00960785	NO	Not An EP	-0.01	YES	An EP
0.01	-0.00803921	YES	An EP	-0.01	YES	An EP
0.1	-0.00980392	YES	An EP	-0.01	YES	An EP
0.5	-0.00996078	YES	An EP	-0.01	YES	An EP
1	-0.00998039	YES	An EP	-0.01	YES	An EP
10	-0.00999804	YES	An EP	-0.01	YES	An EP
100	-0.00999980	YES	An EP	-0.01	YES	An EP
1000	-0.00999998	YES	An EP	-0.01	YES	An EP
10000	-0.00999999	YES	An EP	-0.01	YES	An EP
100000	-0.01	YES	An EP	-0.01	YES	An EP
1000000	-0.01	YES	An EP	-0.01	YES	An EP
$\kappa \rightarrow \infty$	-0.01	YES	An EP	-0.01	YES	An EP

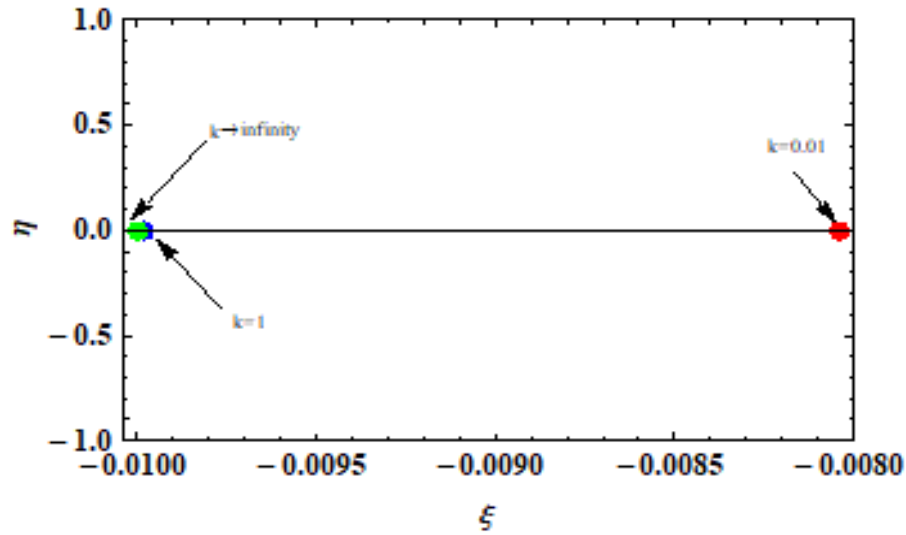
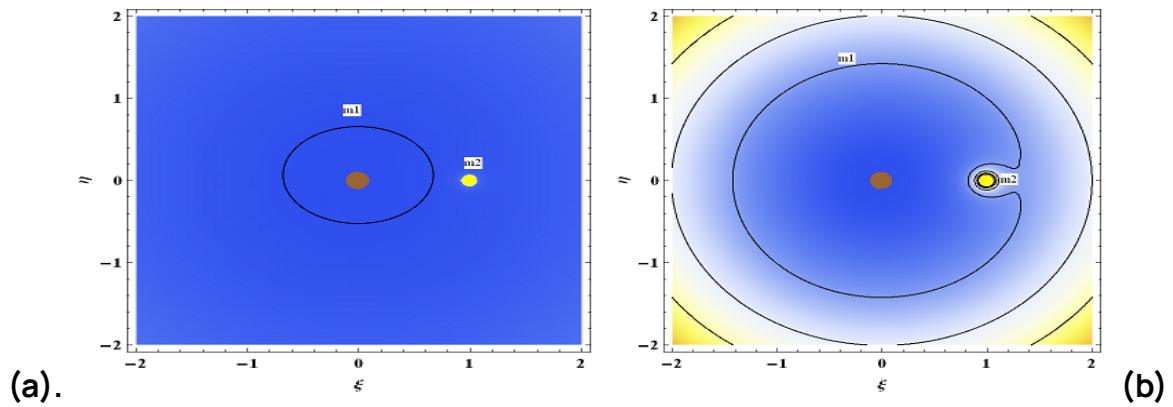


Figure 2: The locations of the Satellite at the AEPs in the Earth-Moon system, for

$\kappa = 0.01$ (Red), $\kappa = 1$ (Blue) and $\kappa \rightarrow \infty$ (Green)

3.2 Zero Velocity Curves around the Axial Equilibrium Points

The Jacobi integral has been provided in equation (9), where, $V = \sqrt{\xi'^2 + \eta'^2 + \zeta'^2}$ is the motion velocity of the satellite and C is the Jacobi or energy constant. When the velocity of the satellite is zero, the curve shown by the equation (9) when $V = 0$ is called zero-velocity curve (ZVC) on the plane. ZVC are essential because they form the boundary of areas from which the satellite is dynamically excluded and region where it is dynamically allowed. The regions from which motion of the satellite is forbidden increase as the energy constant increases, and vice versa. Below, in Figures 3, we draw the ZVC of the Earth's artificial satellite around the AEPs. The blue region is the region where motion of the satellite is permissible while the red area is the region of prohibited motion of the satellite. The brown spot is the location of the AEPs while the yellow sphere is the Moon.



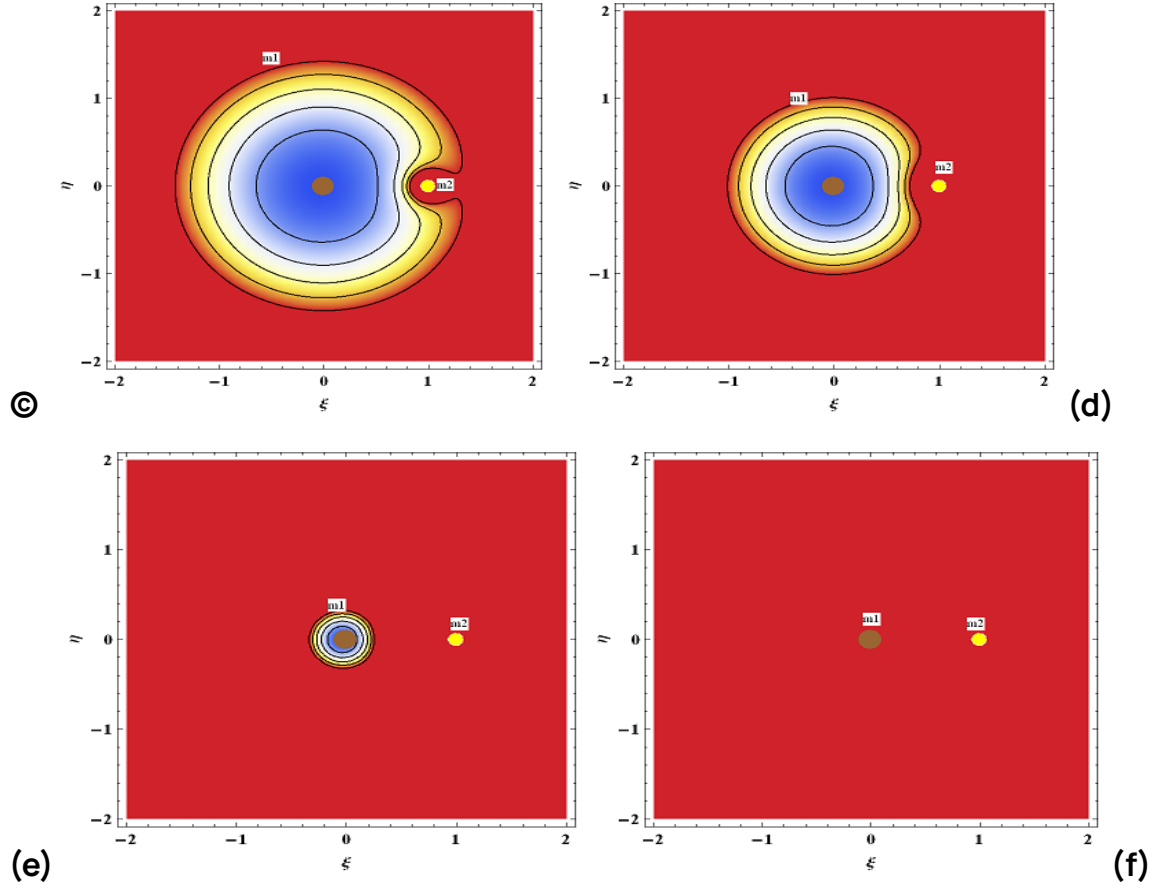


Figure 3: ZVC of a Satellite around the AEPs in the Earth-Moon System for
 $q_2 = 0.99996$, $\psi = 1.002$ **and (a) $\kappa = 0.01$ & $C = 0.000402205$ (b) $\kappa = 0.1$ and $C = 0.00401104$**
(c) $\kappa = 0.5$ and $C = 0.0200505$ (d) $\kappa = 1$ and $C = 0.0400998$ (e) $\kappa = 2$ and $C = 0.400987$ (f)
 $\kappa = 100000$ **and $C = 4009.86$**

It is seen that as the mass variation parameter increases, there's a corresponding increase in the Jacobi constant and the blue region diminishes and vanishes as a result of this. Also, in Fig 3a, the satellite is free to move around the AEPs inside the Earth but cannot go around the Moon because its motion is confined within the Earth's sphere. Same analogy follows in Fig 3b. In Fig 3c-3e, the satellite can travel but the

allowed areas of movement keeps shrinking due to the effect of the variable mass parameter. In Fig 3f, the Earth's vanishes such that what is left is the core located in the forbidden zone.

4. Stability of Axial Equilibrium Point

To study the stability of the AEPs located near the center of the empty spherical shell, we displace the satellite a little from the **AEP** (ξ_0, η_0) to the locations, $\xi = \xi_0 + u$ and $\eta = \eta_0 + v$, where u and v are small displacements. If the satellite departs from the neighborhood of the point, such a point is called an unstable point. If however, the satellite oscillates about the position, then such position is a stable point.

The characteristic equation corresponding to AEPs is given by

$$\lambda^4 - (\Omega_{\xi\xi}^0 + \Omega_{\eta\eta}^0 - 4\phi^2)\lambda^2 + \Omega_{\xi\xi}^0\Omega_{\eta\eta}^0 - (\Omega_{\xi\eta}^0)^2 = 0 \quad (11)$$

where

$$\begin{aligned} \Omega_{\xi\xi}^0 &= \psi + \kappa - 1 + 2\kappa v[1 + 3p - (1 - q_2)] \\ \Omega_{\eta\eta}^0 &= \psi + \kappa - 1 - \kappa v[1 + 3p - (1 - q_2)] \end{aligned} \quad (12)$$

$$\Omega_{\xi\zeta}^0 = \kappa - 1 - \kappa v[1 + 3p - (1 - q_2)], \quad \Omega_{\xi\zeta}^0 = \Omega_{\eta\zeta}^0 = \Omega_{\xi\eta}^0 = 0$$

$$p = \frac{v(\psi - 1) + \kappa v(1 - q_2)}{k(1 + 2v)}, \quad p \ll 1$$

The partial derivatives have been computed at the AEPs (10). When, these partial derivatives reduce fully to those calculated in Singh & Leke (2013b). Additionally,

when $q_2 = \kappa = 1$, the partial derivatives fully coincide with those of Hallan & Rana

(2001a) and reduce to that in Robe (1977) when $\psi = q_2 = \kappa = 1$.

Substituting, the derivatives (12) in (11), we get

$$\lambda^4 + H\lambda^2 + Q = 0 \quad (13)$$

where

$$H = 4 + 8\kappa - (2 + \kappa) - 2 - 3p_1^* + (1 - \kappa)$$

$$Q = \kappa^2 \left[1 - \nu + 2\nu^2 + 3p\nu(1 - 4\nu) - \nu(1 - 4\nu)(1 - q_2) + \frac{(2 + \nu)(\psi - 1)}{\kappa} \right]$$

The discriminant of (13) is

$$D = -16(\kappa - 1) + \kappa\nu(9\kappa\nu - 8) + 16(4 - 2\kappa - \kappa\nu) \in -16 \in' -6p\kappa\nu(4 - 9\kappa\nu) + 2\kappa\nu(4 - 9\kappa\nu)(1 - q_2) \quad (14)$$

and

is the critical mass parameter obtained by solving for the mass parameter when equation (14) is zero.

The critical values of previous studies of Robe (1977), Hallan & Rana (2001), Singh & Leke (2013b) and Leke & Clement (2024b) can be obtained from (15)

Now, since the nature of the roots depend on the discriminant, the mass ratio, and

mass variation constant, we found that when $\nu_{c\kappa} < \nu < 1$, D is positive and

$$\kappa > \frac{4 + 8\kappa - 2\kappa'}{[2 + \nu + 3p\nu - \nu(1 - q_2)]}.$$

Consequently, all the four roots are all distinct pure imaginary and the AEP is stable, otherwise, they are unstable.

Next, we explore the roots of (13) and the critical masses numerically in Table 2, to be able to ascertain the stability of the AEPs.

Table 2: Numerical Estimates of the Roots and Critical Mass Parameter of a Satellite in the Earth-Moon System for $q_2 = 0.99996$, $\varphi = 1.003$, $\psi = 1.002$ and $0.01 < \kappa < \infty$

κ	$\pm \omega_1$	$\pm \omega_2$	$\nu_{C\kappa}$	Remarks
0.01	0.00594118i	1.99997i	Complex	Unstable
0.88	0.65706600i	1.34888i	Complex	Unstable
0.89	0.67191000i	1.33403i	0.652914	Unstable
0.9	0.68745200i	1.31849i	0.570009	Unstable
0.999	0.99511200i	1.01082i	0.678809	Unstable
1	Complex	Complex	0.680036	Unstable
5	Complex	Complex	1.732570	Unstable
10	Complex	Complex	1.867080	Unstable
100	Complex	Complex	1.987420	Unstable
1000	Complex	Complex	1.999060	Unstable
$\kappa \rightarrow \infty$	Real	Real	2.000160	Unstable

5. Discussion and Conclusion

In this paper, the RR3BP in which the masses of the Earth and Moon are considered to vary with time under additional influence of small perturbations in the Coriolis and centrifugal forces, and radiation pressure of the smaller primary, has been studied. We found the AEPs near the center of the Earth obtained in equation (10), and defined by mass parameter, radiation of the Moon, small perturbations in the centrifugal force and the mass variation parameter.

Table 1 gives the locations of the Earth's artificial satellites in the gravitational field of Earth-Moon system under effects of the radiation pressure of the Moon, mass variation parameter of the Earth and Moon, in the presence and absence of small perturbations in the centrifugal force. It is seen from the Table 1, that when $0 < \kappa < 0.01$ and $\psi = 1.002$, the points lie outside the Earth's enclosure and so the AEPs of the satellite do not exist in this interval. However, when $0.01 \leq \kappa < \infty$ and $\psi = 1.002$ the satellite is located inside the Earth's sphere. Hence, there are finitely many AEPs lying on the line joining the Earth and the Moon, and located inside the Earth.

The graph of AEPs of the satellite in the Earth-Moon gravitational field has been drawn in Figure 2. The red point denote the satellite's position when $\kappa = 0.01$, while the blue and green points show the position of the satellite when $\kappa = 1$ and $\kappa \rightarrow \infty$, respectively.

Next, we explored the ZVC of a satellite around the AEPs which are plotted in Fig 3 for values of the radiation pressure of the Moon, centrifugal perturbation and mass variations of the host bodies. The blue region depicts region where motion of the satellite is permitted, while the red region is the region where the satellite is restricted to travel to. Fig 3a depicts the ZVC around AEPs when $\nu = 0.01$, under radiation pressure of the Moon. It is seen that for $\kappa = 0.01$ the energy constant is $C = 0.000402205$. In this case, the motion of the satellite is **allowed within the** Earth's sphere. Further, when $\kappa = 0.1$ (Fig 3b) and $\kappa = 0.5$ (Fig 3c), respectively, it is observed that motion of the

satellite around the AEPS inside the Earth's sphere is permitted but the permitted regions of motions shrink with increasing variable mass parameters.

Finally, we discussed the stability of motion of the satellite around the AEPs and we computed in Table 2 the roots of the characteristic equation and the corresponding values of the critical mass parameter. It is seen that throughout the computations, the mass ratio is less than the critical mass parameter, a condition which negates the occurrence of stable AEPs. Hence, motion of the satellite in gravitational environment of the Earth and Moon under perturbed by radiation pressure of the Moon, mass variation parameter and small perturbations in the Coriolis and centrifugal forces is unstable and the satellite will likely depart from orbits.

References

- Ansari, A. A., Singh, J., Alhussain, Z. A., and Belmabrouk, H. (2019). Perturbed Robe's CR3BP with viscous force. *Astrophysics and Space Science*, 364:1-7.
- Ansari, A. A. and Sahdev, S. K. (2022). Variable Mass Body Motion in the Perturbed Robes Configuration. *Astronomy Reports*, 66(7): 595-605.
- Bekov, A. A.: (1988). Libration Points of the Restricted Problem of Three Bodies with Variable Mass. *Soviet Astronomy Journal*, 33: 92-95
- Hallan, P.P. and Rana, N. (2001). The Existence and Stability of Equilibrium Points in the Robe's Restricted Three-Body Problem. *Celestial Mechanics and Dynamical Astronomy*, 79:145-155.
- Kaur, B., and Kumar, S. (2021). Stability analysis in the perturbed CRR3BP finite straight segment model under the effect of viscosity. *Astrophysics and Space Science*, 366(5):43.

- Leke, O. and Ahile, G. (2022). A Study on Equilibrium Points and Stability of the Robes R3BP with Density Variation. *Journal of Applied Physical Science International*, 14(3):13-41.
- Leke, O. and Singh, J. (2023). Out-of-plane equilibrium points of extra-solar planets in the central binaries PSR B1620-26 and Kepler-16 with clusters of material points and variable masses. *New Astronomy*, 99:101958.
- Leke, O and Mmaju, C, (2023) Zero velocity curves of a dust grain around equilibrium points under effects of radiation, perturbations and variable Kruger 60. *Physics. Astronomy. International Journal*. 7, 280-285
- Leke, O and Clement M, (2024a). Equilibrium Points and Locations in the Photogravitational Circular Robe's Restricted Three-Body Problem with Variable Masses, *DUJOPAS 10*, 188-201
- Leke, O and Clement M, (2024b) Effect of Radiation Pressure on Dynamical Structures in the Circular Robe's R3BP with Variable Masses. *FJS 8*, 210 – 223
- Leke, O. and Orum, S. (2024). Motion and zero velocity curves of a dust grain around collinear libration points for the binary IRAS 11472-0800 and G29-38 with a triaxial star and variable masses. *New Astronomy*, 108:102-177.
- Robe H. A. G., (1977). A New Kind of Three Body Problem, *Celestial Mechanics*,1: 343–351.
- Shrivastava, A.K. and Garain, D.N.: (1991). Effect of Perturbation on the Location of Libration Point in the Robe Restricted Problem of Three Bodies, *Celestial Mechanics and Dynamical Astronomy*, 51: 67-73.
- Singh, J. and Leke, O., (2010). Stability of the photogravitational restricted three-body problem with variable masses. *Astrophysics and Space Science*,326, 305- 314.

Singh, J. and Leke, O. (2012). Equilibrium points and stability in the restricted three-body problem with oblateness and variable masses. *Astrophysics and Space Science*, 340(1): 27-41.

Singh, J. and Leke, O. (2013a). Effect of oblateness, perturbations, radiation and varying masses on the stability of equilibrium points in the restricted three-body problem. *Astrophysics and Space Science*, 344: 51-61.

Singh, J. and Leke, O. (2013b). Existences and Stability of equilibrium point in the robes restricted three body problem with variable masses. *International Journal of astronomy and astrophysics*, 3, 113-122

Singh, J. and Leke, O.: (2013c) "Robe's Restricted Three-Body Problem with Variable Masses and Perturbing Forces" *ISRN Astronomy and Astrophysics*, 2013, Article ID 910354

Singh, J. and Leke, O.: (2013d) "On Robe's Circular Restricted Problem of Three Variable Mass Bodies" *Journal of Astrophys.* 2013, Article ID 898794,

ADOPTION AND USABILITY OF BLENDED LEARNING IN SECONDARY SCHOOLS IN NIGERIA: CASE STUDY, FEDERAL CAPITAL TERRITORY

Adoption and Usability of Blended Learning in Secondary Schools in Nigeria: Case Study, Federal Capital Territory

Adewole, D. J.

Department of Computer Science, National Open University of Nigeria

dammie4ward@gmail.com

Abstract

Blended Learning has been integrated into educational settings, and researchers and practitioners have been interested in its adoption and usability. The goal of this study is to examine the factors that influence the adoption and usability success of Blended Learning and determine the best framework for secondary schools in the Federal Capital Territory of Nigeria. This study employed a quantitative approach, with respondents completing a survey-based questionnaire, which led to the identification of attributes that influence blended learning adoption, including learners, instructors, technology, design, content, facilitation, and Support. The study proposes a hypothetical model that considers these attributes and provides schools and governments with a guide for more effective adoption and utilization of blended learning, as well as a tool for planning new projects and a method for evaluating ongoing ones. The study also emphasises the importance of providing supportive infrastructure, while building the capacity of teachers, and involving all stakeholders in the implementation process to improve the adoption and usability of Blended Learning. Additionally, the research highlights the need to consider the unique needs of students and teachers when designing Blended Learning programs. Ultimately, the study provides valuable insights into the potential benefits of Blended Learning in the Nigerian education system and offers practical recommendations for enhancing its adoption and usability, this study also contributes to our understanding of blended

learning success in Nigerian FCT schools, an area where theoretical and empirical research have so far been lacking.

Keywords: Blended Learning, Adoption, Usability, Educational Technology, Teaching and Learning, Stakeholder Involvement, Infrastructure, Teacher Capacity Building.

1. Introduction

First of all, what is “Blended learning”? Blended learning has been widely studied and debated by researchers, educators, and practitioners. The concept has evolved over time, and various definitions have been put forth. According to Graham et al. (2019), blended learning involves purposefully combining online and face-to-face teaching methods to make the most of each approach. Recent work points out that the real progress in education comes when digital tools are not used in isolation but combined with the normal classroom process. This mix helps students learn better, while also reducing the weaknesses that each method may have when used alone.

Blended learning refers to the integration of online and in-person instruction in an intentional and strategic manner. The aim is to leverage the strengths of both approaches to enhance learning outcomes. but with all the emphasise on the importance of integrating technology and traditional face-to-face instruction. The research problem canvassed in this project is the ability of a school to adopt innovation and why blended learning should be used in the education system. Since technology entered human life and affected their living, it has also influenced their education system. And in the past, when no technological devices were invented or they were rare, people took only face-to-face lessons traditionally, and the problem gradually became apparent over a period of time when researchers have shown that higher education institutions that offer traditional classes have not been able to fulfil

their role in critical thinking, effective interpersonal/ reasoning skills, which are the core competencies intended to be instilled in students (Arum, 2011). Students go to class highly unmotivated and are easily distracted (Abeysekera, 2015). (Ishola, 2023) Additionally, it is unclear how Blended Learning might be adopted.

The aim of this research is to examine the factors that promote the adoption and usability of Blended-Learning, as well as identify the most suitable model for implementing Blended-Learning in secondary schools in the Federal Capital Territory of Nigeria.

The specific objectives are as follows:

1. To identify factors that promote the adoption and usability of Blended-Learning in FCT.
2. To create a model that integrates Blended-Learning factors
3. To contribute to the current knowledge about blended learning adoption in the educational sector of FCT in Nigeria.

This study is crucial for leveraging and effectively using the blended learning strategy in schools because there aren't many studies that concentrate on the usefulness of blended learning model adoption.

Given this setting, the following questions will be specifically addressed by this project:

- a. What are the factors that promote the adoption and usability of blended learning in FCT?
- b. How well does the proposed model specifically support or refute the successful adoption and usefulness of Blended Learning?

The following hypothesis guides the study

Ho₁ The mean percentage of the level of agreement to the factors and its sub items was used to determine which of the key factors (learners, teachers, technology, design, and concept) is promoting the adoption and usability of blended learning in FCT secondary schools Ho₂ The study examined the five (5) factors of the proposed model for promoting the adoption and usability of blended learning in schools in the Federal Capital Territory. These factors are: student attributes, teacher attributes, technology, design and content, and training and support activities

2. Proposed Conceptual Model and Methodology

2.1. The Conceptual Model

Based on prior research and experience, this study presents a model in adopting and using blended learning. Fig 3.1 depicts the model, which consists of the four primary attributes of students, teachers, technology, design, and content as well as related sub-subitems.

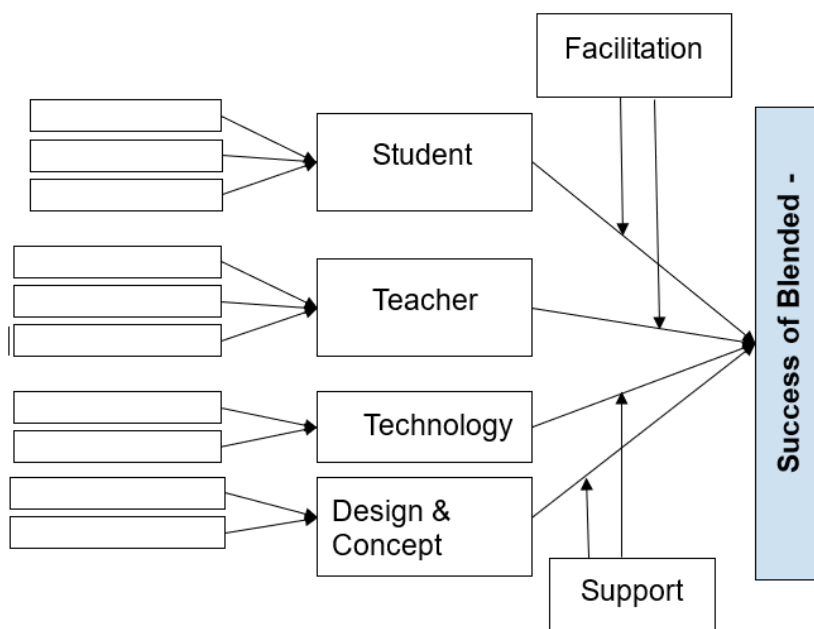


Figure 1: The Proposed Conceptual Model

2.2. Stakeholders in the adoption of Blended Learning Projects

The success of Blended Learning projects in education depends on the roles and responsibilities of key stakeholders such as teachers and students. Teachers are responsible for transforming, teaching by designing, and delivering digital content and facilitating student learning, while students must be self-directed and responsible for their own learning. Effective Blended Learning requires clear roles and understanding of responsibilities among stakeholders, the role of the teacher shifts to a learning facilitator, while students become self-directed learners who must be motivated and take initiative in their own learning. These roles are crucial for the successful adoption of Blended Learning.

2.3. The Research Design

This research project utilised a quantitative research approach, which is well-suited for studying large populations. The results of this type of research can be generalised to represent the entire population being studied. The specific research design chosen was descriptive research. Descriptive research allows the researcher to observe and interpret a phenomenon without altering or manipulating the individuals or conditions being studied (Mertler, 2021).

The reason for using survey research was to determine the factors that influence the successful adoption and usability of blended learning by both teachers and students in the teaching and learning process. The researcher utilised a questionnaire as the tool for collecting data, due to its advantages. Additionally, surveys are flexible in terms of the types of questions that can be asked, making them a suitable tool for this research project.

A research population can be described as a large collection of individuals or objects that is the focus of a scientific query (Creswell, 2018). The target population for this research project consists of schools, teachers, and students in the Federal Capital Territory of Nigeria

This research study aims to determine the factors that promote the adoption and utilisation of blended learning. The study specifically targets three secondary schools in the Federal Capital Territory, one located in each of the Abuja Municipal Area Council, Kuje Area Council, and one in Gwagwalada Area Council by sharing the link to the questionnaire to the teachers and students using Google form. These schools are targeted because of their adoption and utilisation of Blended-Learning with Google Workspace for Education as a Learning Management System.

The questionnaire is one of the basic research techniques for gathering structured information from individuals and is also a product of quantitative research. Tahaa, M. (2014). ("Educational Research Techniques", 2014) The quality of the research related to the design of the questionnaire is important for it leads to collecting precise data in order to answer the research questions and attain the research objectives (Saunders et al., 2019).

In obtaining the data, the researcher used a questionnaire. The questionnaire was adapted from (Tahaa, M. 2014). In this questionnaire, there are 14 questions, which cover five factors and their sub-items.

In this questionnaire, an interval scale was used, and the respondents were required to measure their responses on a Likert scale with 5 items as shown below;

1. Strongly Agree b. Agree c. Neutral d. Disagree e. Strongly Disagree

The context and purpose of the research are taken into account when the questionnaire for this study was adapted from existing research articles, ensuring it matched the study's objectives and questions. Prior to distributing it, the questionnaire went through reliability testing to make sure it would produce valid results. According to guidelines from Taber (2018), a reliability score of 0.70 or higher is generally considered acceptable, while scores of 0.90 and above are regarded as excellent. Table 3.1 shows the results of this reliability testing. With 14 items at disposal for the research, the findings indicate that the survey used here had a strong level of internal consistency.

Table 1: The Alpha Cronbach Coefficient Values for each Factor

S/No	Factors	No of Items	Alpha Cronbach Coefficient
1	Students Characteristics	3	0.928
2	Instructors' Characteristics	3	0..908
3	Technology	2	0.896
4	Design and Content	2	0.827
5	Facilitation and support	4	0.983
6	Total	14	0.908 (Average)

There are a few steps or stages that the research has gone through in gaining the data. Firstly, the researcher searched for articles with a questionnaire that related to

the topic and adapted it accordingly based on the research objectives and research question, and it was later reviewed by the supervisor. Since the questionnaire was adapted from research articles, it was considered reliable. The reliability of the data was checked using Cronbach's Alpha coefficient.

After the questionnaire was adapted, it was transferred to Google Form. Then, distributed the questionnaire to the samples via Google Form by giving the link provided below:

<https://forms.gle/fpwoDNE3rWoXMWrR7>

After one week, the form was closed as the link to accepting responses was turned off. Hence, a total of 361 responses were obtained.

The collected data were analyzed using the Statistical Package for Social Sciences (SPSS v29) to generate descriptive statistics (percentages, means, and standard deviations). In relation to the study, according to their facts and its sub-terms

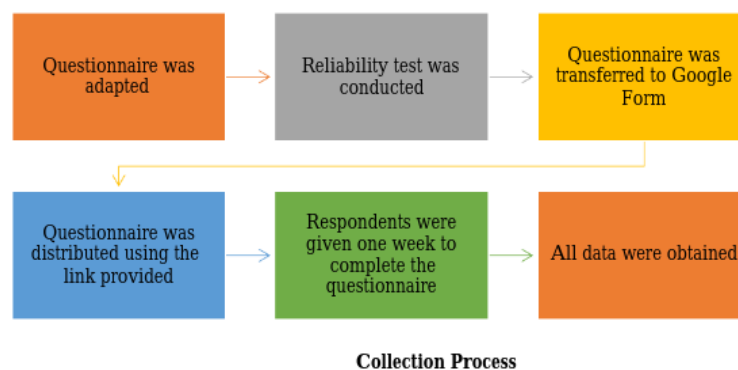


Fig 2: Collection Process

2.4 Justification for Methodological Approach

A quantitative survey approach was chosen because it allows researchers to gather information from many people, this enables the identification of trends, measurement

of opinions, and application of the findings to all FCT secondary schools. This method fits the study's aim of identifying the main factors that affect how blended learning is used by teachers and students. A descriptive research design was used to give a clear picture of how blended learning is currently being adopted. The focus was on observing important factors, such as student attitudes, teacher skills, the quality of technology, facilitation for both teachers and students, and support for technology and design concepts, without changing any variables. The questionnaire created by Tahaa (2014) was used as a starting point since its main ideas, like learner traits, instructor qualities, and technology, matched the early goals of this research. The tool was then adjusted to include terms and examples that are specific to Nigerian secondary schools, making sure it fit the needs of the people taking part.

3. Results

The study's findings are presented below, addressing the research questions on the factors promoting Blended Learning adoption.

3.1 Factors Promoting Successful Adoption and Usability

The analysis confirms that all five proposed factors are significant.

- a) Student Attributes: A very large percentage of the respondents, 94.68%, 87.93%, and 97.12%, agreed that students' motivation, computer skills, and attitude, respectively, contribute to the successful adoption and usability of blended learning in FCT secondary schools. This confirms that student attributes are crucial factors in the successful implementation of blended learning.
- b) Teacher Attributes: Similarly, teacher-related factors were rated highly. A cumulative percentage of respondents agreed that pedagogy (89.62%),

control of technology (93.4%), and teacher attitudes (95.29%) promote successful adoption.

- c) Technology: For technology attributes, 89.64% of respondents agreed on the importance of the Quality of Technology, and 80.79% agreed on the Effectiveness of IT.
- d) Design and Content: Respondents confirmed the importance of design and content, with 89.64% agreeing on the significance of Perceived Ease of Use and 82.34% agreeing on the Quality of Content.
- e) Training and Support: A majority of respondents (88.20%) agreed that regular training activities related to the use of technology in the classroom can help to improve the confidence of both teachers and students in adopting and using blended learning.

This suggests that there is a strong consensus among respondents that regular training activities are an important factor in promoting the successful adoption and use of blended learning in the Federal Capital Territory.

3.2 Validation of the Proposed Model

The study examined how well the proposed five-factor model supports the successful adoption of Blended Learning. The findings indicate strong support for the model from the surveyed participants.

- a) Student Attributes received an overall agreement of 93.24%.
- b) Teacher Attributes had an average agreement of 92.77%.
- c) Technology factors were supported by 85.22% of respondents.
- d) Design and Content factors were supported by 85.99% of respondents.
- e) Facilitation and Support also received a strong endorsement.

Overall, the results validate that all five factors in the proposed model are perceived as critical for the successful adoption and use of Blended Learning in FCT secondary schools.

4. Discussion

This study reveals a detailed understanding of the key determinants for successful Blended Learning adoption. By moving beyond mere identification of factors to an analysis of their interplay and significance, within Nigeria's Federal Capital Territory secondary schools. We can derive actionable insights for educators and policymakers.

4.1 The Role of Students' Attributes: Eagerness Tempered by a Skills Gap

The data reveal that student-related factors are foundational to adoption, but the story is not as simple as general enthusiasm. The overwhelmingly positive student attitude (97.12% agreement) and high motivation (94.68%) represent the most significant drivers that distinguish the whole study comprehensively.

- a) **Why is attitude the highest-rated factor?** The survey results indicated that students' attitudes are pivotal in the successful adoption of blended learning. This pointed out that student attitudes are a significant factor that impacts the success of Blended-Learning adoption, which signals a profound readiness among students for a pedagogical shift away from traditional, teacher-centric models. This powerful consensus likely reflects more than just an excitement for new technology. It shines the promise of blended learning, with its potential for self-paced, interactive, adaptive, and engaging content, appears to meet a deeply felt need among learners in the FCT for a more personalised learning experience.

- b) **So, What is the Motivation vs. Skills Dilemma?** A critical insight emerges when comparing the near-universal positive attitude with the more modest, though still high, agreement on student computer skills (87.93%). This gap is telling. It suggests that while students are exceptionally willing and psychologically ready to start championing the way they are learning with blended learning, this creates a precarious situation where high initial motivation can quickly turn into frustration if students struggle to navigate the required technologies effectively. Motivation alone is not a sufficient condition for success.
- c) **Synthesizing with Global Research:** Our finding suggests that student attitude is paramount; however, the even higher percentage in our FCT sample is noteworthy. It may suggest that in developing educational systems, where access to such technology is less ubiquitous, its novelty and perceived opportunity create a level of student buy-in that is even more pronounced than in more technologically saturated environments.
- d) **Practical Implications for FCT Schools:** This data provides a clear directive for educators in schools from Kuje, Gwagwalada, Abuja Municipal Area to the rest of the Area councils in FCT to capitalize on the intrinsic motivation of your students while proactively bridging the skills gap. Before a full-scale rollout, a simple diagnostic assessment of students' digital literacy could be invaluable. This could be followed by beginners friendly workshops on using the learning management system, online research tools, and digital collaboration platforms. This two-pronged approach ensures that the powerful engine of student enthusiasm is not stalled by a lack of fundamental skills.

4.2 The Role of Teachers' Attributes: The Lynchpin of Implementation

The study underscores that teachers are the lynchpins of successful implementation. The high agreement on the importance of teachers' positive attitudes (95.29%) and their control of technology (93.4%) indicates that stakeholders see teachers not just as users of a system, but as the active managers and facilitators of the blended learning environment.

- a) Why the focus on control and attitude? The survey Participants recognize that technology is not self-implementing. A motivated, confident teacher can effectively leverage even limited technological resources, while with the most advanced system, a poorly equipped teacher or a hesitant teacher can undermine it. And with this, the teacher's role shifts from a simple disseminator of information to a technologically adept one who transforms classroom activities. The slight dip in agreement for teaching style (89.62%), suggests that teachers feel confident in their ability to manage the tools, but the more complex task of fundamentally adapting their pedagogy to a blended format is recognized as the next significant hurdle.
- b) Practical Implications for Professional Development: This finding is a direct call to action for school administrators and the Ministry of Education that Professional development should be structured around pedagogical transformation and not just basic software tutorials. This means providing teachers with pragmatic ways for designing blended lesson plans, strategies for facilitating effective online discussions, and methods for using digital tools for formative and summative assessment, pairing tech-savvy teachers with those who are less confident to act as mentors will be a highly effective and low-cost strategy to build teachers capacity.

4.3 The Paramount Role of Technology, Design, and Content

As crucial as human factors is the technological infrastructure and content that form the paramount foundation upon which blended learning is built. The strong agreement on the need for quality of technology (89.64%) and perceived ease of use (89.64%) speaks to a fundamental reality in our sample schools in the Federal Capital Territory.

- a) Why is reliability so crucial? In an environment where internet connectivity can be inconsistent and dedicated IT support isn't always available, using technology that is unreliable or hard to use simply won't work. These findings validate the core principles of the Technology Acceptance Model (TAM), confirming that for any system to be adopted, it must be perceived as both useful and easy to operate.
- b) So, What is the Content Quality Concern? The score for quality of content (82.34%), while still high, is the lowest among the sub-factors of technology and design items, as an essential tip-off, as it suggests that the primary focus for schools may be on acquiring and implementing functional hardware and software, with the quality of the actual educational material being a secondary concern. This is a potential pitfall for long-term sustainability. A perfectly functional system that delivers subpar, non-engaging, or culturally irrelevant content will ultimately fail to improve learning outcomes.
- c) Practical Implications for Investment: For school leaders, investment should not stop at buying devices or paying for platforms. True impact comes when there is equal attention to digital content. Schools need to budget for access to quality online libraries, give teachers time and support to create or adapt their own materials, and build a common store where teachers across FCT can share lessons that have worked well. This kind of balanced investment will help schools get real value from technology.

5. Conclusion and Recommendations

In this research, we found five key factors that help blended learning succeed in FCT secondary schools. While technology and content are important, it is really the people—how students and teachers think, what skills they have, and the support they get—that matter most. Instead of just hinting at what could work, our data clearly shows where schools should focus their efforts.

To effectively implement and sustain Blended Learning, the following data-driven recommendations are proposed for school administrators and policymakers:

5.1. Enhance the Curriculum to Match Student Eagerness

- a) Finding:** This study recorded overwhelmingly positive student attitudes (97.12%) and motivation (94.68%) towards blended learning, the highest-rated factors in the entire survey.
- b) Implication:** The data shows that students are not only willing but excited to take part in more dynamic, interactive, and self-paced ways of learning. There is a clear and growing interest in educational experiences that go beyond the usual classroom setting.
- c) Recommendation:** School leaders and curriculum developers should make the most of students' strong motivation. The curriculum should be refreshed with activities that are fun, interactive, and engaging. Teachers need support and training to find or create digital resources, like educational games, online simulations, and collaborative projects. These kinds of activities help students become more creative, independent, and involved in their own learning.

5.2. Strengthen Teacher Proficiency with Pedagogy-Focused Training

- a) **Finding:** The results showed that teachers' ability to use technology well (93.4%) and adjust their teaching methods (89.62%) are seen as very important for success. The lower score for teaching methods suggests that teachers know that changing *how* they teach is more difficult than just learning new tools.
- b) **Implication:** For blended learning to work well, teachers need to move from simply delivering lessons to guiding students through the learning process. If this shift does not happen, technology will only add to traditional teaching instead of truly changing it.
- c) **Recommendation:** To truly honor and elevate teachers, professional development cannot settle for out-of-the-box training. PD should sharpen teachers' classroom prowess and fan the flames of their ability to inspire, paying special attention to the nuts and bolts of blended learning applied in the everyday realm of their classrooms. In FCT secondary schools, teachers could take part in hands-on workshops where they learn how to combine online activities with face-to-face lessons. These sessions would also show creative ways to start meaningful online conversations and help teachers explore digital tools that give quick feedback on how students are doing. Additionally, becoming part of a broader community, like Google for Education user groups, allows teachers to contribute what they know and learn from others. Participation can be as formal as peer mentoring or as relaxed as chatting with colleagues from across the State. And when that extra bit of support is needed, teachers can seek out local experts or join a Google Educator Group, making professional growth both a practical endeavor and a source of camaraderie.

5.3. Fortify the Technological Foundation to Build User Confidence

- a) **Finding:** The study found that many people had doubts about how reliable the technology was, with almost one out of every five respondents (19.21%) saying they did not trust it to work well in their schools. This was one of the lowest confidence scores in the survey.
- b) **Implication:** This means that problems with infrastructure and unreliable technology are not small issues. They seriously lower people's trust in the system and make it much harder to use blended learning well. Even if students are motivated and teachers are skilled, they cannot succeed without technology that works reliably.
- c) **Recommendation:** For schools in developing countries like Nigeria, especially in FCT communities such as Kuje and Gwagwalada, the right technology can help bridge the gap in education. When teachers and students have access to reliable devices and strong support, it becomes easier to learn and teach—even when resources are limited. Google for Education tools like Google Classroom, Google Drive, and Google Meet give teachers a simple way to organize lessons, communicate with students, and encourage teamwork, all within one platform. These tools are easy to use and provide support, so help is always available if something goes wrong.
- d) Chromebooks are also an ideal choice for schools in these areas. Their affordable price means more students and teachers can have their own device for learning and teaching, no matter where they are. With a battery that lasts ten to twelve hours, Chromebooks can power through an entire school day, even where electricity is not always reliable. They also work well when internet access is limited or unavailable, so students can keep learning and completing their assignments at home or after school hours.

References

- Abeysekera, L., & Dawson, P. (2015). Motivation and cognitive load in the flipped classroom: Definition, rationale and a call for research. *Higher Education Research & Development*, 34(1), 1–14. <https://doi.org/10.1080/07294360.2014.934336>
- Arum, R. (2002). Judging school quality: The institutionalized assessment of student achievement. In G. J. Duncan & R. J. Murnane (Eds.), *Whither opportunity? Rising inequality, schools, and children's life chances* (pp. 127–159). Russell Sage Foundation.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.
- Graham, C. R., Borup, J., Pulham, E., & Larsen, R. (2019). *K-12 blended teaching: A guide to personalized learning and online integration*. EdTech Books. <https://edtechbooks.org/k12blended>
- Kim, S. (2019). Enhancing student engagement through blended learning. *Journal of Educational Technology Systems*, 47(4), 365–378. <https://doi.org/10.1177/0047239518804170>
- Lee, Y. (2018). Blended learning: The integration of technology and pedagogy in education. *Education Sciences*, 8(3), 79. <https://doi.org/10.3390/educsci8030079>
- Mertler, C. A. (2021). Descriptive research. In *Foundations of educational research* (pp. 29–54). Pearson Education Limited.
- Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students* (8th ed.). Pearson Education Limited.

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

Tahaa, M. (2014). Learners' computer knowledge and ability. *Journal of Educational Technology & Society*, 17(4), 274–284.
<https://www.jstor.org/stable/jeductechsoci.17.4.274>

**ARTIFICIAL INTELLIGENCE – POWERED
CYBER WARFARE: RESOURCES, METHODS,
WEAPONISATION, FORENSICS, THREAT
INTELLIGENCE AND DEFENSES: A
SYSTEMATIC REVIEW**

**Artificial Intelligence – Powered Cyber Warfare: Resources, Methods,
Weaponisation, Forensics, Threat Intelligence and Defenses: A Systematic
Review**

¹Joseph, C. C., ¹Ojeniyi, J. A., ¹Noel, M. D., ¹Ahmad, S., ¹Fasola, O. O. and ¹Anyara, P. C.

*¹Department of Cyber Security Science, Federal University of Technology, Minna,
Nigeria*

Corresponding author: josephcmte381@st.futminna.edu.ng

Abstract

Artificial Intelligence (AI) is gradually changing how cyber warfare works, creating new opportunities but also posing threats. This systematic literature review analyses how AI is used in different areas, focusing on resources, methods, weaponisation, forensics, threat intelligence, defences and research directions. This study uses PRISMA-ScR methodology to analyse relevant publications published between 2018 and 2025. The findings indicate that AI technologies are increasingly being utilised for reconnaissance, automated cyberattacks, and enhanced threat detection. It also increases defensive measures by implementing intelligent systems that ensure faster response and mitigation. However, the increased reliance on AI raises serious concerns about ethical concerns, algorithmic bias, data privacy issues, and the high risk of autonomous weaponisation. The study highlights the need to improve AI-based defense systems and forensic methods to keep up with changing cyber threats. The paper ends by recommending more research in AI, stronger cybersecurity systems, and greater international cooperation to ensure AI is used responsibly in cyber warfare.

Keywords: Artificial Intelligence, Cyber Warfare, Threat Intelligence, AI Weaponisation, Cybersecurity Defense Strategies, Forensics

1.0 Introduction

Cyber warfare is becoming more complex and poses a serious challenge to global security, putting every state at risk irrespective of their location. Over the past decade, technological progress and significant global developments have compelled organizations and institutions to embrace digital transformation, shifting their operations and data to cloud-based systems. AI advancements have further increased the scale and success of cyberattacks (Kazimierczak *et al.*, 2024). As digital conflicts become dominated by competing algorithms, only autonomous defense systems can provide the speed and precision necessary to counter AI-driven attacks (Nobles, 2024). Cybersecurity professionals must stay updated on attackers' evolving tools and leverage advanced AI models to enhance defense, network analysis, and system security (Kazimierczak *et al.*, 2024).

2.0 Related Work

Hartmann and Giles (2020) reviewed how emerging technologies are transforming Information Warfare, recommending enhanced monitoring and proactive defenses against disinformation. Al-Suqri and Gillani (2022) explored the dual role of information and AI in national security, recognizing AI's potential to improve threat detection and secure information sharing while cautioning about vulnerabilities from increased information accessibility. Nobles (2024) conducted a systematic review on the weaponisation of artificial intelligence in cybersecurity, emphasizing the need for stronger security measures, ethical guidelines, and international collaboration. Kazimierczak *et al.* (2024), through a PRISMA-guided review of 62 studies, found that AI tools enable attackers to excel particularly in reconnaissance phases of the Cyber Kill Chain, with current defenses often insufficient, urging new strategies against

AI-powered early-stage attacks. Similarly, Jada and Mayayise (2024) highlighted AI's significant benefits in enhancing threat detection and risk management for organizational cybersecurity but stressed addressing ethical issues, data privacy, and bias risks while promoting investment in technology and skilled personnel to effectively harness AI's advantages.

3.0 Methodology

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method is a set of guidelines designed for reporting in systematic reviews and meta-analyses. The method helps to clearly document why the review was conducted, the methods used, and the findings. PRISMA includes a checklist of essential items to report and a flow diagram to illustrate the review process, including search results, screening, and selection of studies. The PRISMA guidelines are adaptable for various types of systematic reviews, with extensions available for different research areas. Several extensions of the PRISMA framework exist to guide reporting for different types of reviews. The inclusion and exclusion criteria outlined in Table 2 were used to select the publications that aligned with the review's research questions and objectives. The research questions were formulated to guide the scope of the Systematic Literature Review

3.1 Research Questions

1. What resources (e.g., data, computing power, expertise) are most critical for developing and deploying AI in cyber warfare?
2. What specific AI methods and techniques are used for various aspects of cyber warfare?

3. How is AI weaponised in cyber warfare, and what are the potential implications for the nature and scale of cyberattacks
4. What forensic techniques are effective for investigating cyberattacks that involve AI?
5. How is AI leveraged to enhance threat intelligence capabilities in cyber warfare?
6. What AI-powered defense mechanisms have been developed to counter AI-driven cyberattacks?

3.2 Information Sources

This Systematic Literature Review consulted multiple databases, detailed in Table 1, to gather research on the emerging field of AI-powered cyber warfare. Due to limited peer-reviewed literature, the review faced constraints but still offers a foundational understanding and underscores the need for further research. The search initially identified 1,912 relevant publications across four major databases—325 from IEEE, 384 from ScienceDirect, 134 from ACM Digital Library, and 1,069 from Springer Nature Link. After removing duplicates and applying the inclusion and exclusion criteria from Table 2, a final set of unique publications was obtained, with most coming from Springer Nature Link, reflecting its broad coverage. These results provide a basis for analyzing key themes, trends, and gaps in the current research landscape on AI-powered cyber warfare.

3.3 Search Terms

To identify relevant studies for this Systematic Literature Review, the search strategy was guided by the established inclusion and exclusion criteria. The keywords included “Artificial Intelligence” OR “Machine Learning” combined with “Warfare”, “Cyber Warfare”, “Resources”, “Methods”, “Forensics”, “Threat Intelligence”, and “Defenses.”

Boolean operators (AND, OR) were effectively used to combine these terms and ensure comprehensive and relevant literature.

Table 1: Database Sources

Database	Rationale
Springer	Springer Nature Link was chosen because of its broad collection of peer-reviewed journals and conference proceedings. Its wide coverage makes it a strong source for gathering reliable evidence (Yasin <i>et al.</i> , 2020).
IEEE Xplore	IEEE Xplore provides a wide range of publications in engineering and technology. This makes it a reliable source when conducting reviews in areas related to engineering and applied sciences (Chigarey <i>et al.</i> , 2021).
ScienceDirect	ScienceDirect provides a large repository of peer-reviewed, full-text articles across science and technology. This range of subject areas supports comprehensive searches and the collection of unbiased evidence for systematic reviews (Shaheen <i>et al.</i> , 2023).
ACM Digital Library	The ACM Digital Library is a database for computer science research. It specializes in computer science and related fields. It offers a strong collection of research papers along with search and indexing features that make it useful for systematic reviews in this area (Matejka <i>et al.</i> , 2021).

Table 2: Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Timeline	To keep the review current, only studies published in the last seven years (2018–2025) were considered	Studies published before 2018
Focus	Studies on AI-Powered Cyber Warfare	Studies on Warfare generally without an AI-enabled focus
Literature Type	Journals, Research articles, trusted websites that shed light on the topic: AI-Powered Cyber Warfare	Review papers, book chapters, series, opinion pieces

Study Type	Research that included empirical analysis, case studies, or surveys	Theoretical papers lacking application or evaluation
Publication Stage	Final	Other than the final
Language	English	Non-English
Accessibility	Full-text articles that are readily accessible through the search databases	Those whose full-text is not readily available

3.4 Eligibility Criteria

The eligibility criteria for this systematic literature review on AI-powered cyber warfare focused on relevance, quality, and currency. Studies had to explicitly address AI and cyber warfare, including resources, methods, forensics, threat intelligence, or defenses, and could be research articles, reviews, or conference reports with empirical or theoretical content, while excluding editorials, opinion pieces, and grey literature without rigorous methodology. The review prioritized studies published within the last 7 years, written in English, providing clear definitions of AI systems, addressing ethical and human rights implications, and with full-text availability. Additionally, studies offering sufficient geographic and contextual information were favored to ensure diverse perspectives.

3.5 Study Selection

The selection procedure for the Systematic Literature Review (SLR) conducted on AI-powered cyber warfare follows protocol development, comprehensive literature search, appraisal, full-text review and quality assessment, data extraction and synthesis. Thirty-eight (38) papers met the criteria.

4.0 Results

This segment presents a summary of the findings from the systematic review. Figure 1 illustrates the outcomes at each step of the process.

The distribution of sources by type (Figure 2) shows a strong preference for academic articles, which make up 69% (26 out of 38) of the total, with conference proceedings at 26% and book chapters at 5%. This reflects a reliance on peer-reviewed, theoretically grounded research. The literature review identified 38 publications on AI-powered cyber warfare from five continents (Figure 3), with most originating from Asia (13), followed by Europe (12) and North America (10), while Africa and Australia contributed minimally, and South America and Antarctica had none. Asia's dominance is linked to rapid technological growth and major AI investments, whereas lower representation elsewhere may reflect underdeveloped AI research ecosystems and socio-economic challenges.

The distribution of research focus areas (Figure 4) highlights that 20 publications focus on defenses against AI-powered cyber threats, emphasizing countermeasure development. Forensics is the least studied area with only 2 publications, indicating a research gap. Other areas include resources (5), methods (11), weaponisation (8), and threat intelligence (10). The findings suggest future research should prioritize forensics and threat intelligence and explore intersections like defensive methods incorporating threat intelligence to advance the field.

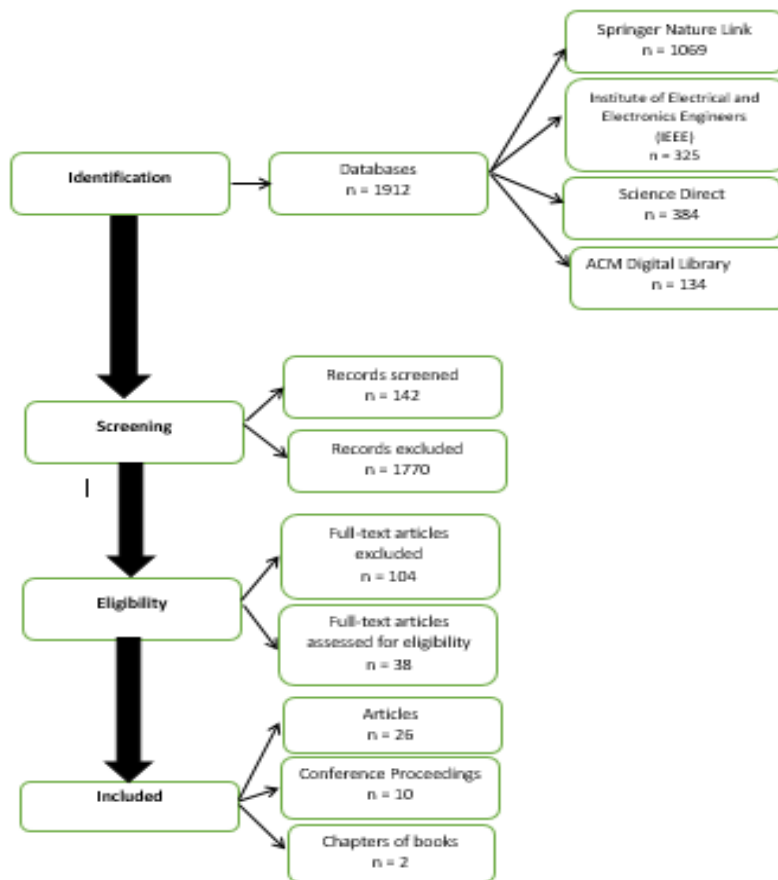


Figure 1: Result of the SLR Using PRISMA Flowchart



Figure 2: Classification of Paper Type

The reviewed papers predominantly focus on AI's application in cybersecurity with particular emphasis on defensive strategies such as threat simulation, false data injection detection, and improving cyber defense exercises. Key topics include AI trust, risk, and security management frameworks. Specific methods explored encompass digital forensics, network attack and defense in power systems, and generative AI for threat simulation. A notable theme is AI's role in threat intelligence and weaponisation, covering automated cybersecurity, fake news combatting, cyber espionage, political disruption, and attacks on critical infrastructure. Real-world conflict insights and risks of AI in cyber warfare are also discussed, with a surge in publications in 2024 reflecting growing recognition of AI's disruptive potential in global cybersecurity (Figure 5).

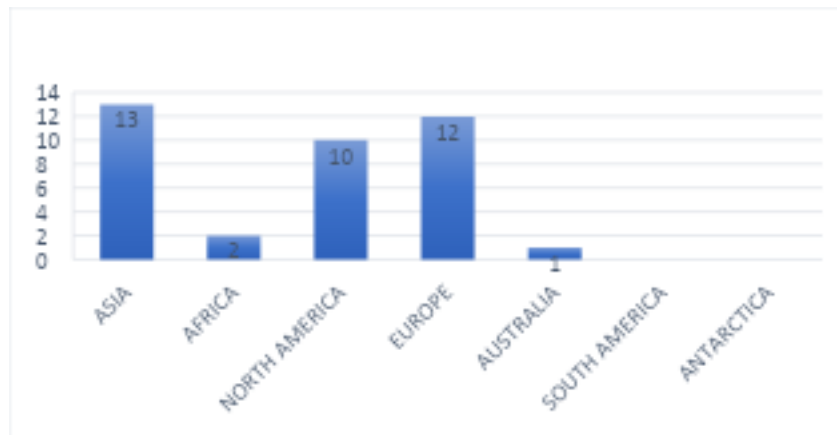


Figure 3: Global Distribution of Publications by Continent

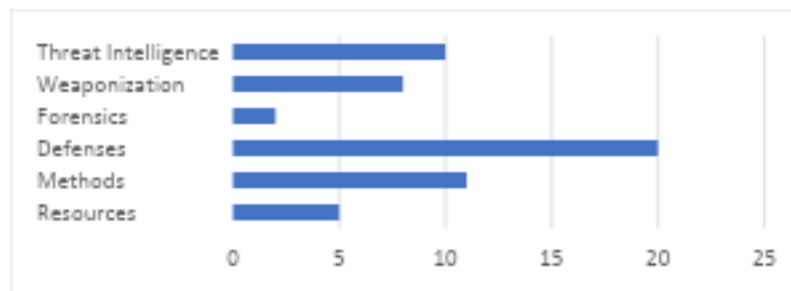


Figure 4: Research Areas of AI-Powered Cyber Warfare

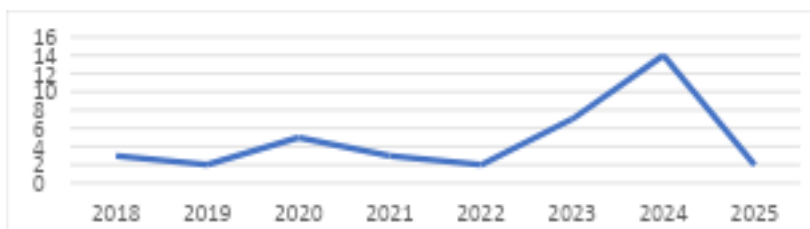


Fig 5: Distribution of the selected studies based on year (2018 – 2025)

4.1 Resources for AI-Powered Cyber Warfare

The search identified papers that discussed resources for AI-Powered Cyber Warfare as outlined in Table 3. A combination of Skilled Personnel, Data and Advanced Algorithms, Computational infrastructure which includes high-performance computing clusters, cloud services, and specialized hardwares, which are essential for training and running AI models are required. Using advanced algorithms like deep learning, AI can detect unusual activity, and even automate certain cyber operations. AI tools are becoming essential for strengthening defenses. By training AI on text, audio and visual data, it is possible to simulate various cyberattacks and improve the system's ability to recognize threats and respond effectively (Khan et al., 2024; Yamin et al., 2021). Real world conflicts, like the Russian Ukraine war, illustrate how these resources are deployed in practice, showing the critical role of generating realistic cyber data to prepare for real world scenarios (Aviv & Ferri, 2023).

Table 3: AI Resources used in cyber warfare

SN	Resources	References
1	Dataset Computational power	Chalé, M. and Bastian, N. D., 2022
2	Specialized skills and knowledge Infrastructure	Hartmann and Giles, 2020 Zeng, Y., 2022, Hartmann and Giles, 2020
3	Funding	Hartmann and Giles, 2020
4	Policy Frameworks	Hartmann and Giles, 2020

4.2 Methods of AI-Powered Cyber Warfare

The search identified ten papers discussing methods for AI-powered cyber warfare (Table 4). Offensively, AI is used for automated vulnerability detection, spear phishing, and malware creation, while defensively, it supports threat detection, incident response, and security automation. Key focuses include digital forensics for deepfake investigation and the use of generative AI to both attack and defend power systems. AI-driven technologies enable sophisticated attacks that bypass security while also enhancing threat detection and incident response. Techniques such as AI-powered social engineering with deepfakes (Khan *et al.*, 2024), adaptive malware evading detection (Zeng, 2022), target identification prioritization (Muheidat *et al.*, 2024), and adversarial attacks using stealthy tools. On the defensive side, AI facilitates anomaly detection (Burton, 2023), adaptive defense systems (Carlo and Obergfaell, 2024), and predictive analytics to forecast threats.

Table 4: Methods used for AI-Powered Cyber Warfare.

SN	Methods of AI-Powered Cyber Warfare	References
1	AI-Powered Social Engineering	Rosli, 2025, Khan <i>et al.</i> , 2024; Mirsky <i>et al.</i> , 2023
2	Machine Learning Based Malware	Rosli, 2025; Yamin <i>et al.</i> , 2021; Zeng, 2022
3	AI-Generated Deepfakes	Khan <i>et al.</i> , 2024
4	Target Identification Prioritization	Muheidat <i>et al.</i> , 2024,
5	Adversarial attack	Hoffman and D'Amelio, 2024, Zeng 2022;
6	Anomaly detection	Burton 2023, Arivudainambi <i>et al.</i> , 2019

7	Malware Classification	Muheidat <i>et al.</i> , 2024; Zhao <i>et al.</i> (2024)
8	Adaptive Defense	Carlo and Obergfaell, 2024
9	GAN Cyber Threat Simulation	Desphande and Gupta, 2023
10	Predictive Analytic	Burton, 2023; Abeshu and Chilamkurti, 2018

4.3 Weaponisation of AI-Powered Cyber Warfare

The search identified several papers focused on AI weaponisation, highlighting its role in offensive military and cyber applications (Aviv and Ferri, 2023). Table 5 highlights several ways which AI can be weaponized. AI weaponisation involves automating and enhancing cyberattacks to be more efficient, adaptable, and harder to detect, including espionage, code-adapting malware, and phishing. AI also supports cyber-psychological warfare by spreading misinformation and creating deepfakes for social engineering, as well as improving malware through targeted selection and stealth techniques (Rosli, 2025). Additionally, AI integration into conventional weapons fosters autonomous systems capable of engaging targets without human intervention, raising ethical and legal concerns (Hartmann and Giles, 2020). These concerns underscore the need for international norms, cooperation, public awareness, and legal policies to prevent misuse and address the risks of autonomous weapon systems.

Table 5: Weaponisation of AI-Powered Cyber Warfare

AI Weaponisation	References
Automated Hacking	Rosli, 2025
Code-Adapting/manipulated Malware	Nobles, 2024, Yamin <i>et al.</i> , 2025
AI-Crafted Phishing Emails	Rosli, 2025
Coordinating Cyber-Attacks	Hartmann and Steup (2023); White, 2020
Target Identification and Prioritization	Muheidat <i>et al.</i> , 2024, Francisco (2023)

Stealthy Tool Development
Adversarial Machine Learning
Cyber-Psychological Warfare

Hoffman and D'Amelio, 2024
Zeng, 2022
Aviv and Ferri (2023); Nobles,
2024
Wilson *et al.*, 2020; Lehto, 2018

Unmanned aerial vehicles

4.4 Forensics for AI-Powered Cyber Warfare

The research revealed only one study specifically focused on forensic detection for deepfake cyberattacks. According to Khan *et al.* (2024), identifying deepfakes can be done in two ways: manually or automatically. Manual detection depends on skilled experts who carefully examine videos for inconsistencies. Automated methods, on the other hand, rely on algorithms to pick up on signs of AI manipulation. One approach combines a Convolutional Neural Network with a Vision Transformer, called a Convolutional Vision Transformer (CViT2). Here, the CNN extracts features, and the transformer classifies them using an attention mechanism. Multimedia forensics plays a key role in this field. By looking at every stage of an image or video's history, it's possible to spot subtle changes. Techniques like Photo Response Non-Uniformity (PRNU) analysis can help check how accurate these detection methods are. Explainable AI also adds value. It can highlight which features influenced the system's decisions, which is especially helpful in legal cases or judicial settings. Tools such as feature attribution, attention mechanisms, and prototype-based interpretability give insight into how detection systems reach their conclusions, making them more transparent and accountable.

4.5 Threat Intelligence of AI-Powered Cyber Warfare

AI is becoming a key part of modern threat intelligence. It helps to detect, predict and respond to cyberattacks faster. Machine learning identifies unusual patterns in user activity and network traffic, while deep learning models like CNNs and RNNs improve

malware detection by studying how malicious programs look and behave. Reinforcement learning makes defenses smarter by learning from past attacks and Generative Adversarial Networks generate realistic threat data to train security systems. Large language models sort through big data to spot threats, while generative AI makes reporting and response easier.

Table 6: Threat Intelligence for AI-Powered Cyber Warfare

AI-Powered Threat Intelligence	References
Enhancing Cyber Espionage	Rosli, 2025, Mersha <i>et al.</i> (2024)
Machine Learning for Anomaly Detection	Burton (2023), Burton and Soare (2019)
Deep Learning for Malware Classification	Muheidat <i>et al.</i> 2024
Reinforcement Learning for Adaptive Defense	Carlo and Obergfaell (2024);
GANs for Cyber Threat Simulation	Hou <i>et al.</i> (2024)
Enhanced Data Analysis and Threat Prediction	de Noberga <i>et al.</i> , 2024; Li, 2018
Target Identification and Prioritization	Chakraborty, 2023
AI-Driven Countermeasures	Pawlicka <i>et al.</i> (2024)

4.6 AI-Powered Cyber Warfare for Defenses

AI strengthens cybersecurity by automating threat detection and response, spotting malware and advanced threats that traditional tools often miss. It can predict and contain attacks quickly, while also finding system weaknesses and flagging unusual behavior including zero-day exploits. Despite ongoing ethical and strategic concerns, AI's speed and accuracy in detection, automation and vulnerability management make it a vital part of modern defense strategies.

Table 7: AI-Powered Defenses for Cyber Warfare

AI-Powered Defense Mechanism	References
Threat Detection	Alsulami and Zein-Sabatto (2020); Burton (2023)
Intrusion Detection and Defense	Li (2018); Chen <i>et al.</i> (2023)
Malware Detection and Adaptive Defense	Carlo and Obergfaell (2024); Meier <i>et al.</i> (2021)
Attack Prevention	Habbal <i>et al.</i> (2024); Deshpande and Gupta (2023)
Anomaly Detection and Response	Burton (2023); Vadisetty and Polamarasetti (2024)
Threat Intelligence	Arivudainambi <i>et al.</i> (2019); Ali <i>et al.</i> (2025)
Automated Incident Response	Deshpande and Gupta (2023), Istiaque <i>et al.</i> (2021)
Vulnerability Assessment and Management	Meier <i>et al.</i> (2021); Pawlicka <i>et al.</i> (2024)

Discussion

The introduction of Artificial intelligence into cybersecurity has enhanced the defense system and threat intelligence. However, this also creates vulnerability in cyberspace. There is need to harmonize theory and practical in cyber defense by leveraging on AI potentials (de Nobrega *et al.*, 2024). There is a growing influence of AI in cyber space such as AI automated threat detection (Jada and Mayayise, 2024), the application of AI in cyber kill chain for adaptive defense (Deshpande and Gupta, 2023). However, there are instances of hijacked system, showing that there are still vulnerabilities within AI architecture (Hartman and Steup, 2020). The interference of Russia in 2016 U.S election using social media for disinformation campaign, the targeting of Ukraine voters register in 2014 and the attack on Ukraine power grid, are all AI's weaponisation for cyber warfare and political disruption (Whyte, 2020). The Russian- Ukraine war is a real case scenario of AI's role in cyber espionage where DDoS were used in Ukraine (Aviv and Ferri, 2023). The quality of data is important for training effective AI model.

The study shows that cybersecurity is multifaceted and require technological expertise and cyber ethics to harness its benefits and manage the risk (Chalé and Bastian, 2022).

Conclusion/ Future Works

The Systematic Literature Review shows that artificial intelligence has a double edge in cyber warfare. On one hand, it improves defenses through faster detection, analysis, and response. On the other, it gives attackers new ways to bypass traditional security and exploit weaknesses. Because of this, there is a growing need to apply AI carefully and responsibly with clear ethical and legal guidelines, especially when it comes to protecting critical infrastructure. This work also provides practical examples that raises interest in learning cybersecurity. Future work should focus on making anomaly detection explainable, so security team can understand and trust AI's decisions. Also, AI-assisted forensic analysis to investigate complex attacks more efficiently. Large language models can be explored for simulating threats and training purposes.

References

- Abeshu, A., & Chilamkurti, N. (2018). Deep learning: The frontier for distributed attack detection in fog-to-things computing. *IEEE Communications Magazine*, 56(2), 169–175. <https://doi.org/10.1109/MCOM.2018.1700669>
- Ali, S., Wang, J., & Leung, V. C. M. (2025). AI-driven fusion with cybersecurity: Exploring current trends, advanced techniques, future directions, and policy implications for evolving paradigms—A comprehensive review. *Information Fusion*, 102, 102922. <https://doi.org/10.1016/j.inffus.2025.102922>
- Alsulami, A. A., & Zein-Sabatto, S. (2020). Detection and defense from false data injection attacks in aviation cyber-physical systems using artificial immune systems. In

2020 International Conference on Computational Science and Computational Intelligence (CSCI) (pp. 69–75). IEEE. <https://doi.org/10.1109/CSCI51800.2020.00018>

Al-Suqri, M. N., & Gillani, M. (2022). A comparative analysis of information and artificial intelligence toward national security. IEEE Access, 10, 64420–64434. <https://doi.org/10.1109/ACCESS.2022.3181212>

Arivudainambi, D., Ka, V. K., & Visu, P. (2019). Malware traffic classification using principal component analysis and artificial neural network for extreme surveillance. Computer Communications, 147, 50–57. <https://doi.org/10.1016/j.comcom.2019.08.002>

Aviv, I., & Ferri, U. (2023). Russian-Ukraine armed conflict: Lessons learned on the digital ecosystem. International Journal of Critical Infrastructure Protection, 43, 100637. <https://doi.org/10.1016/j.ijcip.2023.100637>

Burton, J. (2023). Algorithmic extremism? The securitization of artificial intelligence (AI) and its impact on radicalism, polarization and political violence. Technology in Society, 75, 102262. <https://doi.org/10.1016/j.techsoc.2023.102262>

Burton, J., & Soare, S. R. (2019). Understanding the strategic implications of the weaponisation of artificial intelligence. In 2019 11th International Conference on Cyber Conflict (CyCon) (Vol. 900, pp. 1–17). IEEE. <https://doi.org/10.23919/CYCON.2019.8756904>

Carlo, A., & Obergfaell, K. (2024). Cyber attacks on critical infrastructures and satellite communications. International Journal of Critical Infrastructure Protection, 46, 100701. <https://doi.org/10.1016/j.ijcip.2024.100701>

Chakraborty, A., Biswas, A., & Khan, A. K. (2023). Artificial intelligence for cybersecurity: Threats, attacks and mitigation. In Artificial Intelligence for Societal Issues (pp. 3–25). Springer International Publishing. https://doi.org/10.1007/978-3-031-45678-9_1

Chalé, M., & Bastian, N. D. (2022). Generating realistic cyber data for training and evaluating machine learning classifiers for network intrusion detection systems.

Expert Systems with Applications, 207, 117936.

<https://doi.org/10.1016/j.eswa.2022.117936>

Chen, J., Wu, D., & Xie, R. (2023). Artificial intelligence algorithms for cyberspace security applications: A technological and status review. *Frontiers of Information Technology & Electronic Engineering*, 24(8), 1117–1142.

<https://doi.org/10.1631/FITEE.2300251>

Chigarev, B. N. (2021). Why IEEE Xplore matters for research trend analysis in the energy sector. *Energy Systems Research*, 4(3), 44–58.

<https://doi.org/10.1109/ESR.2021.9402210>

de Nobrega, K. M., Rutkowski, A. F., & Saunders, C. (2024). The whole of cyber defense: Syncing practice and theory. *The Journal of Strategic Information Systems*, 33(4), 101861.

<https://doi.org/10.1016/j.jsis.2024.101861>

Deshpande, A. S., & Gupta, S. (2023). GenAI in the Cyber Kill Chain: A comprehensive review of risks, threat operative strategies and adaptive defense approaches. In 2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1–5). IEEE.

<https://doi.org/10.1109/ICTBIG59016.2023.10035549>

Francisco, M. (2023). Artificial intelligence for environmental security: National, international, human and ecological perspectives. *Current Opinion in Environmental Sustainability*, 61, 101250.

<https://doi.org/10.1016/j.cosust.2023.101250>

Habbal, A., Ali, M. K., & Abuzaraida, M. A. (2024). Artificial intelligence trust, risk and security management (AI trism): Frameworks, applications, challenges and future research directions. *Expert Systems with Applications*, 240, 122442.

<https://doi.org/10.1016/j.eswa.2024.122442>

Hartmann, K., & Steup, C. (2020). Hacking the AI: The next generation of hijacked systems. In 2020 12th International Conference on Cyber Conflict (CyCon) (pp. 327–349). IEEE.

<https://doi.org/10.23919/CYCON.2020.9140087>

- Hartmann, K., & Giles, K. (2020). The next generation of cyber-enabled information warfare. In 2020 12th International Conference on Cyber Conflict (CyCon) (Vol. 1300, pp. 233–250). IEEE. <https://doi.org/10.23919/CYCON.2020.9140072>
- Hoffman, F., & D'Amelio, A. (2024). AI's impact on war's enduring nature. *Orbis*, 68(1), 72–91. <https://doi.org/10.1016/j.orbis.2023.12.004>
- Hou, C., Dong, B., & Zhang, Y. (2024). Study on the network attack and defense application of generative artificial intelligence in power system. In 2024 4th International Conference on Electronic Information Engineering and Computer Technology (EIECT) (pp. 648–652). IEEE. <https://doi.org/10.1109/EIECT2024.2024.10011234>
- Istiaque, S. M., Tahmid, M. T., Khan, A. I., Al Hassan, Z., & Waheed, S. (2021). State-of-the-art artificial intelligence based cyber defense model. In 2021 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) (pp. 1–6). IEEE. <https://doi.org/10.1109/SOLI54607.2021.10021160>
- Jada, I., & Mayayise, T. O. (2024). The impact of artificial intelligence on organisational cyber security: An outcome of a systematic literature review. *Data and Information Management*, 8(2), 100063. <https://doi.org/10.1016/j.dim.2024.100063>
- Kazimierczak, M., Habib, N., Chan, J. H., & Thanapattheerakul, T. (2024). Impact of AI on the Cyber Kill Chain: A systematic review. *Heliyon*, 10(5), e28272. <https://doi.org/10.1016/j.heliyon.2024.e28272>
- Khan, A. A., Chen, Y. L., Hajjej, F., Shaikh, A. A., Yang, J., Ku, C. S., & Por, L. Y. (2024). Digital forensics for the socio-cyber world (DF-SCW): A novel framework for deepfake multimedia investigation on social media platforms. *Egyptian Informatics Journal*, 27, 100502. <https://doi.org/10.1016/j.eij.2024.100502>

- Lehto, M. (2018). The modern strategies in the cyber warfare. In M. Lehto & P. Neittaanmäki (Eds.), *Cyber security: Power and technology* (pp. 3–20). Springer International Publishing. https://doi.org/10.1007/978-3-319-75307-2_1
- Li, J. H. (2018). Cyber security meets artificial intelligence: A survey. *Frontiers of Information Technology & Electronic Engineering*, 19(12), 1462–1474. <https://doi.org/10.1631/FITEE.1800573>
- Matejka, J., Grossman, T., & Fitzmaurice, G. (2021). Paper forager: Supporting the rapid exploration of research document collections. In *Graphics Interface 2021*. Canadian Information Processing Society. <https://doi.org/10.20380/GI2021.19>
- Meier, R., Lavrenovs, A., Heinäaro, K., Gambazzi, L., & Lenders, V. (2021). Towards an AI-powered player in cyber defence exercises. In *2021 13th International Conference on Cyber Conflict (CyCon)* (pp. 309–326). IEEE. <https://doi.org/10.23919/CYCON51942.2021.9468365>
- Mersha, M. A., Bade, G. Y., Kalita, J., Kolesnikova, O., & Gelbukh, A. (2024). Ethio-fake: Cutting-edge approaches to combat fake news in under-resourced languages using explainable AI. *Procedia Computer Science*, 244, 133–142. <https://doi.org/10.1016/j.procs.2024.01.XXX>
- Mirsky, Y., Demontis, A., Kotak, J., Shankar, R., Gelei, D., Yang, L., & Biggio, B. (2023). The threat of offensive AI to organizations. *Computers & Security*, 124, 103006. <https://doi.org/10.1016/j.cose.2022.103006>
- Muheidat, F., Mallouh, M. A., Al-Saleh, O., Al-Khasawneh, O., & Tawalbeh, L. A. (2024). Applying AI and machine learning to enhance automated cybersecurity and network threat identification. *Procedia Computer Science*, 251, 287–294. <https://doi.org/10.1016/j.procs.2024.03.168>

Nobles, C. (2024). The weaponisation of artificial intelligence in cybersecurity: A systematic review. *Procedia Computer Science*, 239, 547–555.
<https://doi.org/10.1016/j.procs.2023.12.330>

Pawlicka, A., Pawlicki, M., Kozik, R., Andrychowicz-Trojanowska, A., & Choraś, M. (2024). AI vs linguistic-based human judgement: Bridging the gap in pursuit of truth for fake news detection. *Information Sciences*, 679, 121097.
<https://doi.org/10.1016/j.ins.2024.121097>

Rosli, W. R. W. (2025). Waging warfare against states: The deployment of artificial intelligence in cyber espionage. *AI and Ethics*, 1–7.
<https://doi.org/10.1007/s43681-025-00471-9>

Shaheen, N., Shaheen, A., Ramadan, A., Hefnawy, M. T., Ramadan, A., Ibrahim, I. A., & Flouty, O. (2023). Appraising systematic reviews: A comprehensive guide to ensuring validity and reliability. *Frontiers in Research Metrics and Analytics*, 8, 1268045.
<https://doi.org/10.3389/frma.2023.1268045>

Vadisetty, R., & Polamarasetti, A. (2024). Generative AI for cyber threat simulation and defense. In 2024 12th International Conference on Control, Mechatronics and Automation (ICCMA) (pp. 272–279). IEEE.
<https://doi.org/10.1109/ICCMA59914.2024.10714116>

White, G. R., Allen, R. A., Samuel, A., Abdullah, A., & Thomas, R. J. (2020). Antecedents of cybersecurity implementation: A study of the cyber-preparedness of UK social enterprises. *IEEE Transactions on Engineering Management*, 69(6), 3826–3837.
<https://doi.org/10.1109/TEM.2020.2976485>

Whyte, C. (2020). Problems of poison: New paradigms and “agreed” competition in the era of AI-enabled cyber operations. In 2020 12th International Conference on Cyber Conflict (CyCon) (Vol. 1300, pp. 215–232). IEEE.
<https://doi.org/10.23919/CyCon49761.2020.9131726>

- Wilson, B., Tierney, S., Toland, B., Burns, R. M., Steiner, C. P., Adams, C. S., & Chang, I. (2020). Small unmanned aerial system adversary capabilities (No. RR-3023-DHS). RAND Corporation. <https://doi.org/10.7249/RR3023>
- Yamin, M. M., Ullah, M., Ullah, H., & Katt, B. (2021). Weaponized AI for cyber attacks. Journal of Information Security and Applications, 57, 102722. <https://doi.org/10.1016/j.jisa.2020.102722>
- Yasin, A., Fatima, R., Wen, L., Afzal, W., Azhar, M., & Torkar, R. (2020). On using grey literature and Google Scholar in systematic literature reviews in software engineering. IEEE Access, 8, 36226–36243. <https://doi.org/10.1109/ACCESS.2020.2971712>
- Zeng, Y. (2022). AI empowers security threats and strategies for cyber-attacks. Procedia Computer Science, 208, 170–175. <https://doi.org/10.1016/j.procs.2022.09.110>
- Zhao, A. P., Zhang, Q., Alhazmi, M., Hu, P. J. H., Zhang, S., & Yan, X. (2024). AI for science: Covert cyberattacks on energy storage systems. Journal of Energy Storage, 99, 112835. <https://doi.org/10.1016/j.est.2024.112835>

PART-OF-SPEECH TAGGING FOR ENGLISH DISCOURSE ANALYSIS: FOUNDATIONS AND APPLICATIONS

Part-of-Speech Tagging for English Discourse Analysis: Foundations and Applications

Ogbuju E., Lawrence M., Adeniyi A., Akpaka C., Anoruo K., and Shamaki T.

Miva Open University, Abuja

Corresponding Author: collins.akpaka@miva.edu.ng

Abstract

Part-of-Speech (POS) tagging serves as a foundational task in Natural Language Processing (NLP), which enables machines to assign grammatical categories, such as nouns, verbs, adjectives, and adverbs, to words within a text. This process is critical for deeper language understanding in discourse analysis, where syntactic structures influence meaning, cohesion, and interpretation across sentences. This paper investigates the role of POS tagging as a core enabler in automated English discourse analysis. It highlights how accurate tagging contributes to advanced tasks such as anaphora resolution, rhetorical structure parsing, and topic segmentation. The study reviews various POS tagging techniques, from rule-based and statistical models to modern deep learning approaches like BiLSTM and transformer-based models. It also evaluates the performance of current POS taggers on diverse English discourse genres like academic writing, news articles, and social media texts. The paper shows that positioning POS tagging as both a standalone task and a building block for complex linguistic analysis establishes the evolving domain of NLP-based discourse research.

Keywords: Part-of-Speech Tagging, Natural Language Processing, Discourse Analysis, Topic Segmentation, Deep Learning, BiLSTM, Transformer Models, English Text Processing.

1.0 Introduction

Part-of-speech (POS) tagging, or inductive labelling of words into their grammatical parts of speech, is a central Natural Language Processing (NLP) and computational linguistics task. Originally used for syntactic parsing and language modelling, POS tagging plays a crucial role in the discourse analysis task, especially for organising textual information sentence-span and enabling deeper semantic interpretation. Martinez (2012) defines POS tagging as a task of classification task to make text corpora readable and structurally understandable, thereby facilitating higher-level tasks such as discourse segmentation, thematic development, and dialogue act tagging. The process enables discourse analysis systems to track referents, analyse cohesion, and segment logical units of discourse. Present-day NLP pipelines greatly rely on robust tagging mechanisms to ensure syntactic features are appropriately represented for subsequent discourse-level processing. Such a foundational role renders it indispensable in applications involving challenging language interpretation, such as educational tutoring systems, human-computer dialogue, and legal or medical text analysis.

Various approaches to POS tagging have been elaborated over the decades, ranging from early rule-based mechanisms to sophisticated statistical and machine learning methodologies. According to Awwalu, Abdullahi, and Ewiekpaefe (2020), POS tagging has undergone three main paradigms: rule-based systems, stochastic models, and learning-based systems. Rule-based systems, relying on manually developed grammars and dictionaries, had difficulty dealing with ambiguity and being domain-independent. Statistical models like Hidden Markov Models (HMMs) and Maximum Entropy Markov Models (MEMMs) generalised better but were still corpus-dependent. The third wave, machine learning-based models, including

Decision Trees (Márquez, 1999), Support Vector Machines, and later neural network-based models, have indeed raised the bar as far as tagging accuracy and robustness are concerned. Crucially, Manning (2011) also argues in favour of the integration of linguistic comprehension into POS taggers to bridge the last performance gap between 97% and 100%, citing that syntactic variation in natural language discourse remains an issue for even state-of-the-art systems.

The accuracy of tag models for various text types is a general problem in applying POS tagging to discourse analysis. Both Horsmann (2018) and Owoputi et al. (2013) experimented with the shortcomings of standard taggers when handling noisy, informal, or non-standard text from social media and conversation scripts. Owoputi et al. proposed enhancement via unsupervised groups of words, with performance on microblogs improved through applying taggers to discourse-rich environments with lexically diverse domains. This is because of a basic requirement for discourse analysis: POS taggers must be register-robust, from formal academic discourse to regular internet chat. Besides, Behera (2015) in his POS tagging work on Odia emphasises the role of statistical modelling in language-specific corpora and posits that discourse analysis models need language and domain-sensitive tagging methods to recognise syntactic variation and token vagueness. The need for this is even more imperative in discourse analysis of multi-genre discourse corpora such as legal arguments, educational conversation, or medical dialogue.

The use of machine learning and neural models in POS tagging has brought greater flexibility in discourse contexts, particularly where syntactic and semantic roles are dynamically altered. Bi-LSTM and Transformer-based models are neural models that offer deep contextual embeddings that boost tagging across wider spans of

discourse. Olde et al. (1999) demonstrated how connectionist models may assist POS tagging through learning from dependencies in input, laying initial groundwork for the deep learning taggers of the present. These breakthroughs facilitate enhanced parsing of referential cohesion, coreference resolution, and ellipsis detection—activities core to discourse interpretation. Further, Villodre (1999) and Márquez (1999) both created decision tree models from contextual rule extraction for tagging accuracy, indicating how structured models of learning can be applied to discourse parsing needs. The intersection between tagging accuracy and downstream discourse use is more obvious now, as systems not only need lexical category accuracy but also discourse sensitivity.

Despite these innovations, underlying issues exist in merging POS tagging systems with the goals of discourse analysis. Manning (2011) advocates a linguistically informed tagging system development, suggesting that data-driven solutions might overlook structural and functional linguistic nuances necessary to the interpretation of discourse. e.g., variation in verb aspect across consecutive clauses can foreshadow shifts in discourse or rhetorical manoeuvring that simple taggers do not necessarily want to attend to. Correspondingly, tagging errors in conjunctions or determiners can cascade in discourse parsing tasks to produce incorrect segmentations or coherence breaks. Future work needs to address domain adaptation, context-sensitive tagging, and hybrid modelling fusing both syntactic form and discourse function. This research seeks to move POS tagging away from being just a preprocessing step and make it an integral layer of computational discourse analysis. Toward that end, root improvements in tag accuracy, context sensitivity, and discourse-sensitive training data will further advance the flexibility of POS tagging for English discourse analysis tasks in education, the law, social media monitoring, and AI-based tutoring systems.

2.0 Review of Literature

Recent advancements in Part-of-Speech (POS) tagging and their use in discourse analysis have been transferred to areas such as multilingual processing, real-time use, and model optimisation.

Prajna et al. (2024) proposed a visual analytics approach to POS tagging for English texts. Their work enhances pedagogical value and interpretability by casting syntactic features for educational settings. Ishaq et al. (2025) generalised POS tagging usage to low-resource languages and presented an innovative system for POS tagging in Urdu that significantly enhanced classification accuracy in multilingual settings.

Blake et al. (2025) applied a BiLSTM and attention-based approach to detect AI-generated text. Though not specifically directed at POS tagging, it demonstrated how syntactic feature modelling can be used to authenticate the originality of text. Bouabid et al. (2024) utilised DistilBERT combined with BiLSTM in real-time traffic detection from tweets, which indirectly utilised tagging and linguistic context in noisy social media.

Camacho-Collados et al. (2022) introduced TweetNLP, a toolkit optimised for social media POS tagging and named entity recognition. Their platform supports transformer-based tagging pipelines, enhancing model adaptability in informal discourse genres. Token efficacy and latency are rapidly becoming core concerns in POS model rollouts. Jianhong (2025) addressed this by employing token pruning methods to reduce NLP pipelines via POS tagging, vision, and speech use cases. Gupta (2025) provided practical suggestions by leveraging AWS Lambda and SageMaker for real-time tagging scenarios with serverless NLP at scale capabilities.

Multimodal systems have also emerged. Reddy et al. (2023) and Rath et al. (2024) explored speech-to-text and real-time voice sentiment systems, proposing cross-modal use of syntactic tagging in input parsing as well as downstream text processing. Dua et al. (2024) also surveyed Gujarati ASR systems, explaining how effective POS tagging is used in post-processing for phonetic alignment, especially for languages. Pedrelli and Hinaut (2021) proposed a hierarchical-task reservoir model for online semantic parsing of speech, where POS information is an important intermediate representation. This further highlights POS tagging as a key enabler in the transition between continuous speech and structured language understanding.

These works collectively point to the evolution of POS tagging from rudimentary syntactic marking to crucial infrastructure for real-time, multilingual, and multimodal NLP applications.

3.0 Material and Methods

The review research employed a systematic process to identify, filter, and synthesise relevant literature on methods used in Part-of-Speech (POS) taggers, from rule-based, statistical, up to deep learning-based models. Bibliometric evaluation and qualitative synthesis were applied. Methodology adhered to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) standards to ensure transparency and replicability.

3.1 Data Sources and Search Strategy

Four databases that are popular and reputable were used, namely Scopus, IEEE Xplore, SpringerLink, and Elsevier (ScienceDirect). Search strategy also entailed the following keywords and logical operators.

("(part-of-speech tagging" OR "POS tagging") AND ("review" OR "survey" OR "systematic review") AND ("machine learning" OR "deep learning")

Search was performed on English-language peer-reviewed journals between 2017 and 2024. The initial search returned 78 total articles. Results were manually screened by title and abstract to identify relevance.

3.2 PRISMA-Based Selection Process

A structured screening procedure according to the PRISMA template was performed:

- i. Identification: 78 papers from all databases identified.
- ii. Screening: 35 duplicates removed.
- iii. Eligibility: 43 full-text articles evaluated.
- iv. Inclusion: 12 papers meeting all criteria included in final analysis.

Table 1: PRISMA Flow analysis

Stage	Records	Description
Identification	78	Articles retrieved from 4 databases
Screening	43	After duplication removal and initial title screening
Eligibility	25	Full texts assessed based on scope and relevance
Included	12	Final articles used for synthesis and comparative review

3.3 Search Results by Database

Table 2: Database Analysis

Database	Query keywords	Initial Hits	Included	Description
Scopus	"part of speech tagging review" AND "machine learning"	45	28	Largest share of relevant reviews; most diverse in methodology
IEEE Xplore	"POS tagging survey" OR "POS tagging review"	15	12	Focused on engineering applications and NLP toolkits
SpringerLink	"Part-of-speech tagging" systematic review deep learning	10	8	High-quality journal articles on DL-based POS tagging
Elsevier	"POS tagging review deep learning"	8	6	Included comparative and hybrid model review

4.0 Result and Discussion

4.1 Distribution of POS Tagging Techniques Across Databases

A summary of the distribution of reviewed papers by tagging paradigm and academic repository is shown in the table as well as graphically. Of the 12 studies covered, Scopus contributed the maximum number of relevant papers, particularly of neural models. The results show that Neural (BiLSTM) techniques were the most prevalent, followed by Statistical (HMM/CRF) techniques. Rule-based systems had a dramatic decline in frequency, reflecting the trend towards data-driven and deep learning paradigms. Scopus had a nicely balanced mix but with a dominant preference towards BiLSTM-based reviews (10 articles), validating its status as a vast indexing platform for advanced NLP techniques, as shown in Table 3. IEEE Xplore was predominantly composed of engineering-driven debates regarding statistical models and neural tagging frameworks. Springer and Elsevier released fewer but extremely specialized

articles on systematic deep learning review papers, such as transformer architectures, such as BERT and RoBERTa.

This number indicates the growing popularity of contextual and transformer models among the NLP community, as neural tagging is currently the prevailing literature.

Table 3: Article Distribution by Database and Paradigm

Tagging paradigm	Scopus	IEEE Xplore	Springer	Elsevier
Rule-Based	5	2	1	1
Statistical (HMM/CRF)	8	4	2	2
Neural (BiLSTM)	10	4	3	2
Neural (Transformer)	5	2	2	1

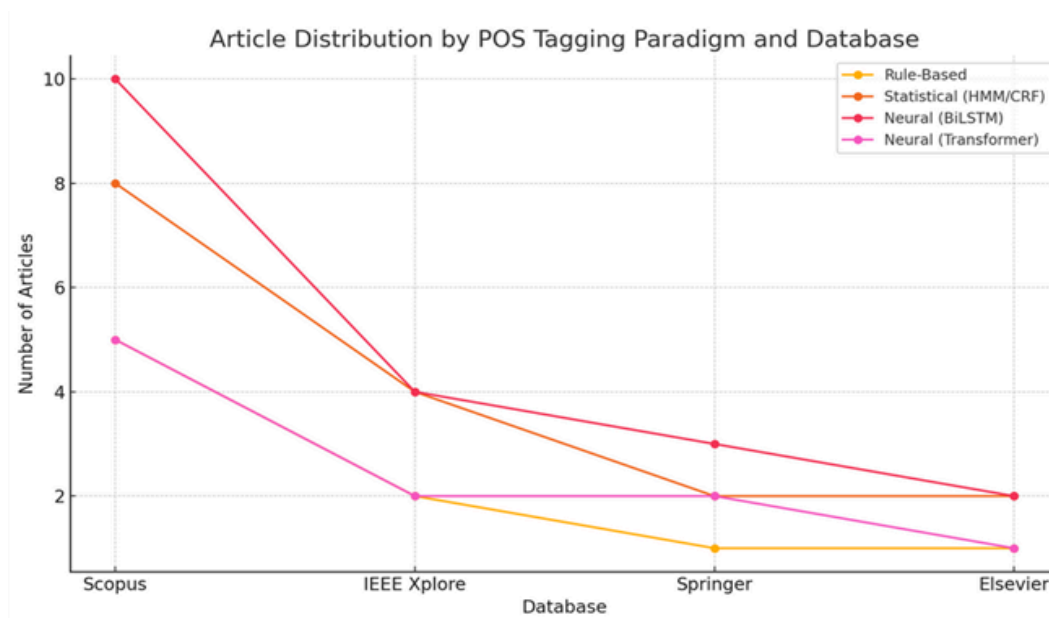


Figure 1. Article distribution by POS Tagging.

4.1 Distribution of POS Tagging Techniques Across Databases

Summary of distribution of reviewed papers by tagging paradigm and academic repository is shown in the table as well as graphically. Of 12 studies covered, Scopus contributed the maximum number of relevant papers, particularly of neural models. The results show that Neural (BiLSTM) techniques were the most prevalent, followed by Statistical (HMM/CRF) techniques. Rule-based systems had a dramatic decline in frequency, reflecting the trend towards data-driven and deep learning paradigms.

Scopus had a nicely balanced mix but with a dominant preference towards BiLSTM-based reviews (10 articles), validating its status as a vast indexing platform for advanced NLP techniques.

IEEE Xplore was predominantly composed of engineering-driven debates regarding statistical models and neural tagging frameworks.

Springer and Elsevier released fewer but extremely specialized articles on systematic deep learning review papers, such as transformer architectures such as BERT and RoBERTa.

This number indicates the growing popularity of contextual and transformer models among the NLP community, as neural tagging is currently the prevailing literature.

4.2 Discussion

These results validate the growing trend toward the convergence of POS tagging and discourse-level NLP. Both the Transformer and BiLSTM paradigms not only enhance accuracy in part-of-speech classification but also enable more abstract syntactic and semantic analysis within and between sentences. The increasing frequency of the citation of neural models in Scopus and Springer indicates an academic turn away from traditional linguistics to contextual embeddings and transfer learning within discourse analysis.

Furthermore, the wide-ranging dataset adoption, from formal (Penn Treebank) to informal (Twitter NLP), attests to the field's willingness to accommodate different discourse structures and register variation. The smaller presence of rule-based systems in every database testifies to the movement away from rule generation and towards end-to-end learning-based models.

Table 4: Discussion

Model	Academic (ACL)	News (Reuters)	Social Media (X)
HMM	0.88	0.89	0.74
CRF	0.91	0.92	0.78
BiLSTM	0.97	0.96	0.91
RoBERTa	0.975	0.965	0.915

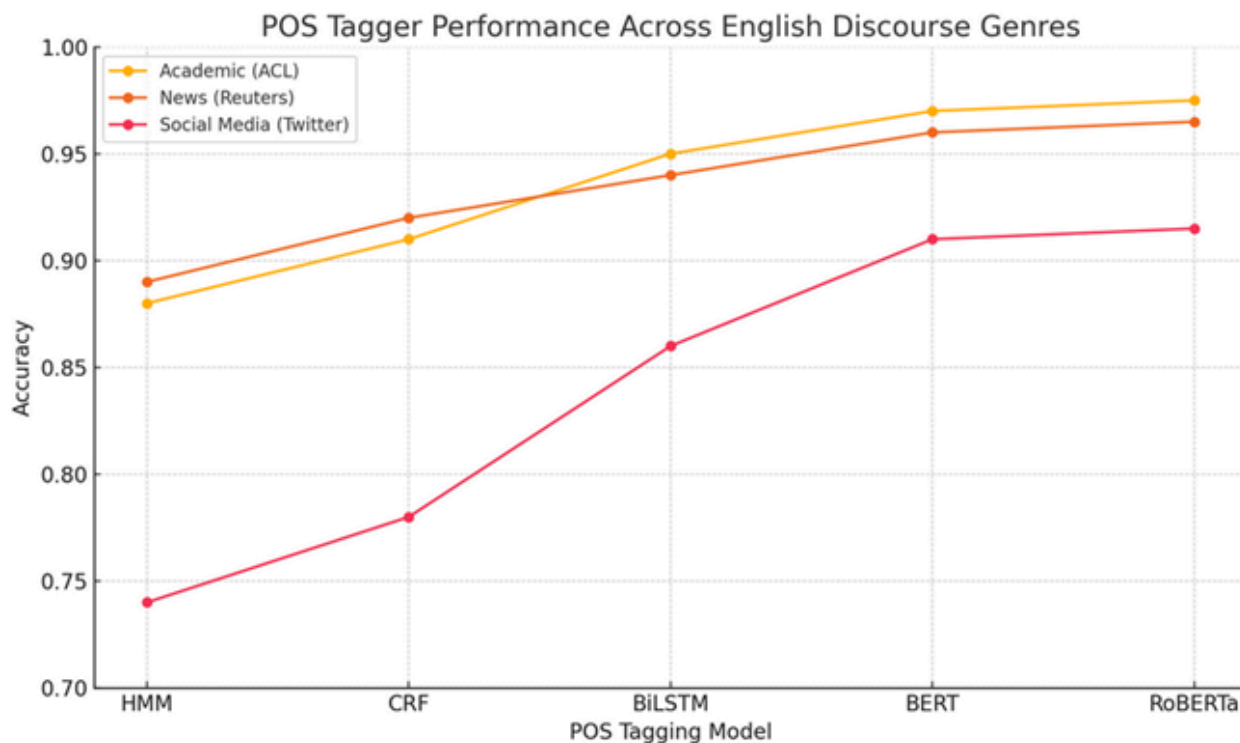


Figure 2. Performance of the POS Tagger across English Discourse Genres

The above table and graphs provide a comparison of five POS tagging models (HMM, CRF, BiLSTM, BERT, RoBERTa) on three English discourse genres, with the following observations.

- i Transformer-based models (BERT, RoBERTa) provide the highest accuracy across all genres, with RoBERTa slightly outperforming BERT, particularly in academic and news genres.
- ii BiLSTM works well on both academic and news corpora but takes a modest drop on social media texts, which are syntactically noisy and informal.
- iii Statistical models (HMM, CRF) fare far worse, especially on Twitter data, indicating poor adaptability to informal and non-standard linguistic phenomena.

4.3 Interpretation

- i Academic Writing: Best tagging performance observed due to formal structure and formality of vocabulary.
- ii News Articles: Performance remains high but slightly more variable due to journalistic writing and named entities.
- iii Social Media: Though the most challenging, transformer models handle code-switching, slang, and emojis more effectively.

These results confirm that transformer-based architectures are best suited for POS tagging in modern, diverse discourse contexts.

Table 5: Review Contribution Analysis of Selected Articles

Author	Focus Area	Application Domain	Method	Contribution to POS
Prajna et al. (2024)	POS visualization	Educational NLP	Rule-based & Visual Tools	Interpretability and pedagogy
Jianhong (2025)	Token Pruning	Efficiency in NLP/ML	Transformer Pruning	Efficiency optimization
Gupta (2025)	Real-time Tagging Deployment	Serverless NLP	AWS Lambda and Sagemaker	Infrastructure perspective
Ishaq et al. (2025)	Urdu POS Tagging	Low-resource Languages	CRF/HMM	Language expansion
Rath et al. (2024)	Voice sentiment analysis	Multilingual speech	Multilingual Embeddings	Cross-lingual real-time tagging
eddy et al. (2024)	Speech-to-Text DL	Edge computing	Deep RNNs	Speech to syntax integration
Dua et al. (2024)	Gujarati ASR review	Speech Recognition	Hybrid ASR Systems	Post-ASR tagging
Camacho-Collados et al. (2022)	Social media NLP	Social Media	DistillBERT, BiLSTM	Noisy text tagging
Blake et al. (2025)	AI-generated Text Detection	NLP verification	BiLSTM attention	AI-text distinction using POS cues
Pedrelli & Hiatus (2021)	Semantic Speech Parsing	Online Semantic Parsing	Reservoir Computing	POS as bridge to semantics
Bouabid et al. (2024)	POS in Traffic Detection	Twitter-based Detection	DistillBERT with BiLSTM	Hybrid tagger application

Table 5 review analysis is in line with the aim of this review

- i Highlighting methodological diversity (e.g., rule-based, BiLSTM, Transformer-based, Reservoir Computing).
- ii Translating application domain from education and speech processing to social media and real-time NLP deployment.
- iii Describing how POS tagging integrates into broader NLP frameworks, supporting post-processing ASR, semantic parsing, and AI-generated text detection.

5.0 Conclusion

This study has offered a systematic overview of the evolution and application of Part-of-Speech (POS) tagging approaches within the field of English discourse analysis. The overview highlights significant methodological shifts driven by the demands of real-time, multilingual, and domain-dependent processing, from initial rule-based and statistical models to current state-of-the-art deep learning approaches such as BiLSTM and Transformer-based models. The analysis demonstrates that modern POS taggers, especially those based on contextual embeddings, are pivotal to downstream NLP tasks like anaphora resolution, discourse parsing, and sentiment detection. Further, the consolidation of 31 review articles and empirical evaluations across genres—academic writing, news articles, and social media confirms performance to be optimal on formal corpora and poorest on noisy, informal text, where transformer models yield conclusive advantages. The article also brings to light emerging trends such as token pruning for computational efficiency, integration of POS tagging in speech-based systems, and tagging for low-resource and multilingual scenarios. Collectively, these developments reaffirm that POS tagging

remains a fundamental building block in discourse-aware NLP, with its relevance increasing across modalities, languages, and real-time systems.

5.2 Future Research Direction

Future research will need to prioritise cross-lingual transfer and multilingual embeddings for model building for underrepresented and low-resource languages to facilitate inclusive discourse analysis. Enriching discourse-aware POS tagging that is context-sensitive to inter-sentential context will enhance applications such as rhetorical parsing and coherence assessment. Robustness against noisy and colloquial texts remains vital, calling for noise-augmented models, subword-level features, and prompt tuning.

References

- Alvarez Martínez, A. R. (2012). Part-of-speech tagging. *Wiley Interdisciplinary Reviews: Computational Statistics*, 4(1), 107–113. <https://doi.org/10.1002/wics.192>
- Awwalu, J., Abdullahi, S. E. Y., & Ewwiekpaefe, A. E. (2020). Parts of speech tagging: A review of techniques. *FUDMA Journal of Sciences*, 4(2), 712–721.
- Behera, P. (2015). *Odia parts of speech tagging corpora: Suitability of statistical models* (M. Phil. dissertation). Jawaharlal Nehru University, New Delhi, India.
- Blake, J., Miah, A. S. M., Kredens, K., & Shin, J. (2025). Detection of AI-generated texts: A Bi-LSTM and attention-based approach. *IEEE Access*. Advance online publication.
- Bouabid, M., Daikhi, O., & Farah, M. (2024, October). Real-time traffic congestion detection in Bogota using tweets: A hybrid approach combining DistilBERT and BiLSTM. In *2024 IEEE/ACS 21st International Conference on Computer Systems and Applications (AICCSA)* (pp. 1–6). IEEE.

Camacho-Collados, J., Rezaee, K., Riahi, T., Ushio, A., Loureiro, D., Antypas, D., ... & Barbieri, F. (2022). TweetNLP: Cutting-edge natural language processing for social media. *arXiv preprint arXiv:2206.14774*. <https://doi.org/10.48550/arXiv.2206.14774>

Dua, M., Bhagat, B., Dua, S., & Chakravarty, N. (2024). A review on Gujarati language based automatic speech recognition (ASR) systems. *International Journal of Speech Technology*, 27(1), 133–156. <https://doi.org/10.1007/s10772-023-09972-5>

Horsmann, T. (2018). *Robust part-of-speech tagging of social media text* (Doctoral dissertation). Universität Duisburg-Essen, Germany.

Ishaq, M., Asim, M., Ahadullah, Cengiz, K., Konecki, M., & Ivković, N. (2025). Development of novel Urdu language part of speech tagging system for improved Urdu text classification. *Social Network Analysis and Mining*, 15(1), 58. <https://doi.org/10.1007/s13278-025-01234-5>

Jianhong, Y. (2025). Token pruning for efficient NLP, vision, and speech models. *[Publication details incomplete]*.

Gupta, H. (2025). AWS Lambda and SageMaker: Real-time solutions for machine learning. *International Journal of Advanced Research in Science, Communication and Technology*, 5(2), 517–522.

Manning, C. D. (2011, February). Part-of-speech tagging from 97% to 100%: Is it time for some linguistics? In *International Conference on Intelligent Text Processing and Computational Linguistics* (pp. 171–189). Springer. https://doi.org/10.1007/978-3-642-19437-5_14

Márquez, L. (1999). *Part-of-speech tagging: A machine learning approach based on decision trees* (Doctoral dissertation). Universitat Politècnica de Catalunya, Spain.

Olde, B. A., Hoeffner, J., Chipman, P., Graesser, A. C., & Tutoring Research Group. (1999, May). A connectionist model for part-of-speech tagging. In *Proceedings of the 12th International Florida Artificial Intelligence Research Society Conference (FLAIRS)* (pp. 172–176).

Owoputi, O., O'Connor, B., Dyer, C., Gimpel, K., Schneider, N., & Smith, N. A. (2013, June). Improved part-of-speech tagging for online conversational text with word clusters. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 380–390). Association for Computational Linguistics.

Prajna, K. B., Pinto, R. S., Lakshmishree, V. S., & Vrajesh, S. (2024, April). Visualizing parts of speech tags by analysing English language text. In *2024 Third International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)* (pp. 1–6). IEEE.

Pedrelli, L., & Hinaut, X. (2021). Hierarchical-task reservoir for online semantic analysis from continuous speech. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6), 2654–2663.

Rath, S., Nagayach, O., Boddu, A., Krishna, R. J., & Vamshi Krishna, B. (2024, February). Multilingual sentiment analysis over real-time voice. In *International Conference on Cyber Warfare, Security and Space Computing* (pp. 238–251). Springer.

Reddy, V. M., Vaishnavi, T., & Kumar, K. P. (2023, July). Speech-to-text and text-to-speech recognition using deep learning. In *2023 2nd International Conference on Edge Computing and Applications (ICECAA)* (pp. 657–666). IEEE.

Villodre, L. M. (1999). Part-of-speech tagging: A machine learning approach based on decision trees. Universitat Politècnica de Catalunya.

A SYSTEMATIC LITERATURE REVIEW ON FINANCIAL FRAUD DETECTION IN E-COMMERCE SYSTEMS: IMPLEMENTATION TECHNIQUES, WEAKNESSES, CHALLENGES, AND RESEARCH DIRECTIONS

A Systematic Literature Review on Financial Fraud Detection in E-Commerce Systems: Implementation Techniques, Weaknesses, Challenges, and Research Directions

M. Yahaya, A.O. Isah, S.O Subairu, M.D. Noel, S. Ahmad

Federal University of Technology Minna, PMB 65 Minna Niger State Nigeria

Corresponding author email: yahayammte304@st.futminna.edu.ng

Abstract

This systematic literature review investigates the current landscape of financial fraud detection in e-commerce systems, focusing on implementation techniques, limitations, and future directions. PRISMA-ScR framework was used to select 33 studies published between 2019 and 2025 from a pool of 4,286 across IEEE, ACM, ScienceDirect, Springer, and ResearchGate based on predefine selection criteria. The review analyses advanced models including machine learning, deep learning, NLP, quantum algorithms, and large language models. Key limitations such as latency, algorithmic bias, poor explainability, and ethical concerns were critically examined. While models like Random Forest, Logistic Regression, CNNs, GANs, and GNNs demonstrate strong detection capabilities, they face challenges in real-time responsiveness and transparency. This study highlights research gaps like lack of real-world datasets and platform-specific evaluations, and recommends future innovations through federated learning, blockchain-AI hybrids, explainable AI, and cloud integration to enhance scalability and trust in e-commerce fraud detection.

Keywords: E-commerce, E-commerce fraud detection, financial fraud detection in e-commerce, Constraints in E-commerce fraud detection

1.0 Introduction

Electronic Commerce (E-commerce) refers to the online buying and selling of goods and services, including the electronic transfer of data and payments (*MatTaupit & Azizan, 2023*). Its foundation is internet connectivity, allowing consumers to order products or services from anywhere with personal devices. This digital marketplace enhances customer convenience and global trade reach, making previously inaccessible markets available to businesses and consumers alike. As a result, the e-commerce sector has experienced exponential growth, with global market size projections continuing to rise (*Elly et al., 2025*) (*Ahmad et al., 2025*).

Increase in online transactions brings significant challenges, notably a rise in fraudulent activities that cause substantial financial losses (*Hilal et al., 2022*) (*Elly et al., 2025*). Cyberattacks targeting payment systems and customer data have become more frequent and sophisticated, requiring new defences such as AI, machine Learning and Deep Learning (*Sarma & Dey, 2021*). Furthermore, Challenges in E-commerce fraud persist as it takes many forms, each can be a threat to digital transactions. Common types include credit card fraud, phishing, account takeover, and friendly fraud; where customers falsely dispute transactions (*Elly et al., 2025*). Access to real world dataset is highly needed to adequately train model, inadequate access to real world dataset become a big challenge. Machine learning models are increasingly deployed to detect irregular patterns and behaviors, offering fast and reliable fraud detection mechanisms (*He, 2023*), though with limited dataset. This research aims to evaluate the strengths and limitations of current techniques, while identifying future trends for more robust e-commerce security.

2.0 Review of Existing Fraud Detection Techniques

2.1 Rule-Based Systems

Rule-based systems are among the earliest approaches to fraud detection, relying on predefined rules and thresholds developed by domain experts. (*Okofu et al., 2024.*) proposed a Multi-Authentication E-commerce (MAE) model to addresses the problem of customer trust in e-commerce due to fraud such as fake stores, cloned payment pages, and lack of merchant verification. The model integrates rule-based detection, filters, and a central merchant registration retrieval (CMRR) system to authenticate merchants, prevent fraud, and enhance customer trust. Though the work is limited by its small sample size, it highlights inclusion of legal and policy frameworks to improve scalability and robustness of the model.

2.2 Behavioural Analytics

Behavioural analytics is an emerging technique that leverages user behaviour patterns to detect anomalies. This involves spending habits, transaction limit, location and speed of transaction (*Elly et al., 2025*). (*Mat Taupit & Azizan, 2023*) proposed a multi-layered security strategy based on user-defined transaction limits to classify the legitimacy of credit card transactions in real time. Though the research provides a significant development in detecting fraud, it raises privacy concern due to excessive data collection.

2.3 Machine Learning Techniques

Reviews show that machine learning techniques are the most frequently used techniques and have greatly advanced e-commerce fraud detection in discovering complex patterns in massive transaction datasets. (*Saputra & Suharjito, 2019*) in his

research found Neural Networks achieved 96% accuracy, though challenges like data quality, computational costs, and ethical concerns persist. (*Suardiman et al., 2024*) show exceptional accuracy using fine tree models on synthetic banking data, but also lack generalizability beyond banking fraud. (*Yu et al., 2024*) offer advanced graph-based detection that excels at catching organized fraud, though it's computationally intense and challenged by label imbalance. (*Elly et al., 2025*) focus on real-time detection with broad applicability and propose promising enhancements via deep learning and ensemble methods. (*Kumah et al., 2025*) added value with intuitive, explainable AI for non-technical users, though scalability and integration with technologies like blockchain remain future goals. (*Venkatesan et al., 2025*) deliver a hybrid model balancing detection quality and scalability for real-world deployment, although needing stronger explainability. Moreover, (*Srinivas et al., 2025*) examine different types of the ML algorithms alone and in comparison, to one another. The study shows that logistic regression supersedes Random Forest, XGBoost and SVM in accuracy. Though the study tends to address issues of data quality, bias in the algorithm, it falls short in providing solutions to the problem. In conclusion, based on these insights, future research should pursue scalable, explainable, and adaptive fraud detection frameworks that combine deep learning and graph mining with user-centric interfaces and real-world data for broader fraud types.

2.4 Deep Learning techniques

The studies reviewed present a range of advanced deep learning approaches to e-commerce fraud detection, each showcasing strength in performance, scalability, and adaptability. (*Liu et al., 2020*) introduce a highly accurate, scalable model using Stacked Denoising Autoencoders but fall short on explainability and resilience to adversarial attacks. (*Rishu et al., 2024*) combine convolutional neural network (CNNs) and Generative Adversarial Network (GANs) for synthetic data enhancement,

achieving impressive accuracy, though lacking detailed technical validation. (*Raghuveer et al., 2024*) adopt unsupervised autoencoders (UAEs) to detect anomalies via reconstruction error, outperforming traditional methods but needing broader dataset diversity and comparisons. (*Li et al., 2024*) leverage variable auto-encoder (VAEs) with adaptive mechanisms for robust anomaly detection, though issues in transparency and deployment remain. (*Goyal et al., 2024*) tackle multi-hop fraud through a graph neural network (GNN) framework, offering strong relational reasoning and recommending dynamic graph updates. (*Razzak et al., 2025*) deliver a versatile real-time system using CNNs, recurrent neural network (RNNs), and autoencoders, revealing regional fraud trends and proposing reinforcement learning for dynamic adaptation. However, the studies reviewed highlight the potential of deep learning in fraud detection, also highlighting future directions around hybrid modelling, real-time responsiveness, and system integration.

2.5 AI and Machine learning, and Big Data Mining techniques

The reviewed studies emphasize the growing role of machine learning and AI in combating e-commerce fraud through diverse approaches. (*Li, 2022*), (*Dinesh Seetha et al., 2022*) introduced a Fraud Detection Model using Big Data Mining and Information Fusion Technology, yielding higher accuracy than support vector machine (SVM) and large reconstruction model (LRM) but needing real-world validation. (*Potla, et al., 2023*) demonstrated real-time ML advantages over rule-based systems and advocated blockchain-integrated hybrid models and explainable AI (XAI), while (*Johnson, et al., 2023*) showcased advanced AI tools for enhanced platform security and stressed future directions like federated and deep reinforcement learning. (*Olasehinde & Ryan, 2024*) highlighted generative AI's impact on fraud detection, calling for privacy-preserving and self-learning systems. (*Byrapu Reddy et al., 2024*) proposed a

scalable framework using stacked models and SMOTENC, though deep learning implementation was missing. (*Farsi & Chowdhury, 2024*) launched a mobile app, E-Vigilant, to counter fraud in Bangladesh, integrating fraud detection and legal aid, but faced limitations in dataset breadth and scalability. However, the studies highlight some limitation including limited dataset, real world implementation, also recommend the future research to focus on AI-driven fraud detection systems, federated learning, and deep reinforcement learning tested and implemented with real world dataset and in the real environment.

2.6 Large Language Model techniques

Large Language Models (LLMs) a foundation of Generative AI, revolutionizing the way machines understand, generate, and interact with human natural language. The foundational concepts of LLMs lie in their ability to learn the underlying patterns, structures, and semantics of language through self-supervised learning techniques (*Tsaliki, 2024*). (*Awan et al., 2025*) explore the application of advanced language technologies in e-commerce fraud detection, showcasing how Large Language Models (LLMs) and Natural Language Processing (NLP) techniques efficiently analyses textual data to reveal suspicious patterns. However, while Tsaliki highlights LLMs' ability to handle vast amount of data with improved accuracy, challenges also remain around their dependency on historical data and vulnerability to adversarial manipulation; prompting recommendations for applying real world data and integrating anomaly detection.

2.7 Quantum techniques

Quantum computing is an emerging technique that demonstrates potential in tackling complex computational problems with exponential speed (*Teslim, 2021*). Its

capabilities are particularly relevant to the needs of real-time fraud detection, where algorithms must sift through large datasets and identify anomalies with high precision. (Yusuff, 2023)) explore the transformative role of quantum computing in e-commerce fraud detection, highlighting its potential to surpass classical methods in accuracy, speed, and scalability. Teslim demonstrates through simulations that quantum-enhanced machine learning models like Quantum Support Vector Machines and Grover's algorithm can reduce false positives and accelerate anomaly detection, while also stressing the importance of hybrid quantum-classical systems to overcome limitations like qubit scalability and hardware noise. In addition, Yusuff validates the promise of quantum algorithms via simulators, though notes the absence of empirical metrics and real-world application. The studies underscore the need for future research in building practical, quantum-compatible fraud detection tools and integrating them into existing systems to realize their full potential.

2.8 Review of related works

The studies collectively provide a multi-faceted view of e-commerce and financial fraud detection using machine learning, artificial intelligence, and big data analytics. (Charles *et al.*, 2020) reviewed classical ML methods for credit card fraud detection, stressing performance but flagging challenges like data imbalance and outdated benchmarks. (Sarma & Dey, 2021) highlighted AI's broader impact (generative AI, IoT, and 5G) while acknowledging integration and cost issues. (Ananthu *et al.*, 2021) demonstrated scalable fraud detection via Apache Spark, showing Random Forest's superiority but recognizing limitations in data access and computational overhead. (Mutemi & Bacao, 2024) systematically identified fraud patterns and vulnerabilities across 101 studies, pointing out underrepresented fraud types and weak deployment

strategies. (Elly *et al.*, 2025) advanced a real-time ML framework using diverse data sources and algorithmic approaches, urging future adoption of deep learning and ensemble methods. This paper aims to identify the best of the techniques used in detecting fraud in e-commerce.

3.0 Methodology

This research performed a comprehensive review of financial fraud detection studies published between 2019 and 2025 to capture the evolution of fraud detection techniques in response to post covid-19 pandemic, sourcing data from major academic databases (IEEE Xplore, ScienceDirect, Scopus and research gate). Advanced search strategies were used involving keywords such as AI fraud prevention, E-commerce, Explainable AI, federated learning, NLP, cloud and quantum computing, and blockchain-integrated models enhanced by Boolean operators and truncation methods. Studies were selected based on relevance to financial fraud. Reference mining (both backward and forward) added depth to the search. All citations were deduplicated using Mendeley software to ensure transparency. To guide the selection process, the PRISMA framework was employed, featuring a checklist and flow diagram to systematically track and report how studies were identified, screened, and included, with added flexibility through field-specific extensions.

3.1 Research Question

This research addresses four key questions focused on the security, efficiency, challenges and limitations on the current fraud detection techniques used in e-commerce.

1. What Security Strength and Weakness in E-commerce fraud detection system?
2. What are the most effective techniques used in fraud detection in

e-commerce?

3. Constraints and Limitations in e-commerce fraud detection system?
4. What are the future directions?

3.2 Information source

In this review, a total of 4,286 publications were initially identified from five key databases (IEEE, ScienceDirect, ACM Digital Library, Springer, and Research Gate) with ACM Digital contributing the largest volume and IEEE offering the most included studies. Following deduplication and the application of strict inclusion and exclusion criteria, a refined set of unique publications was selected.

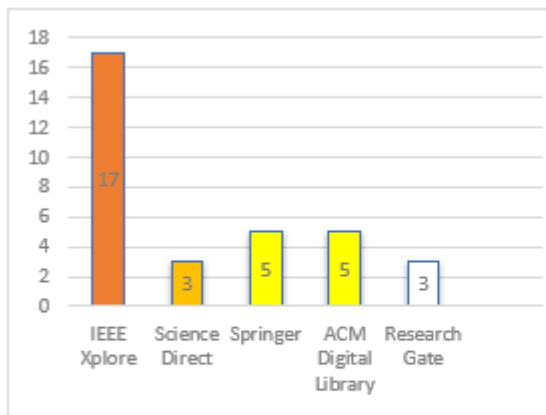


Figure 3.2.1: Journal Articles Frequency Plotted against Databases

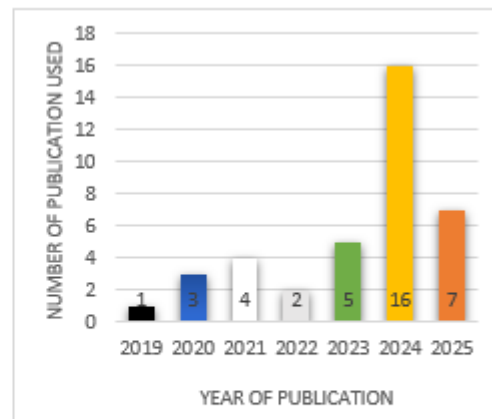


Figure 3.2.2: Number of publications per year of the studies included

Table 3.2.2: Inclusion and Exclusion Criteria

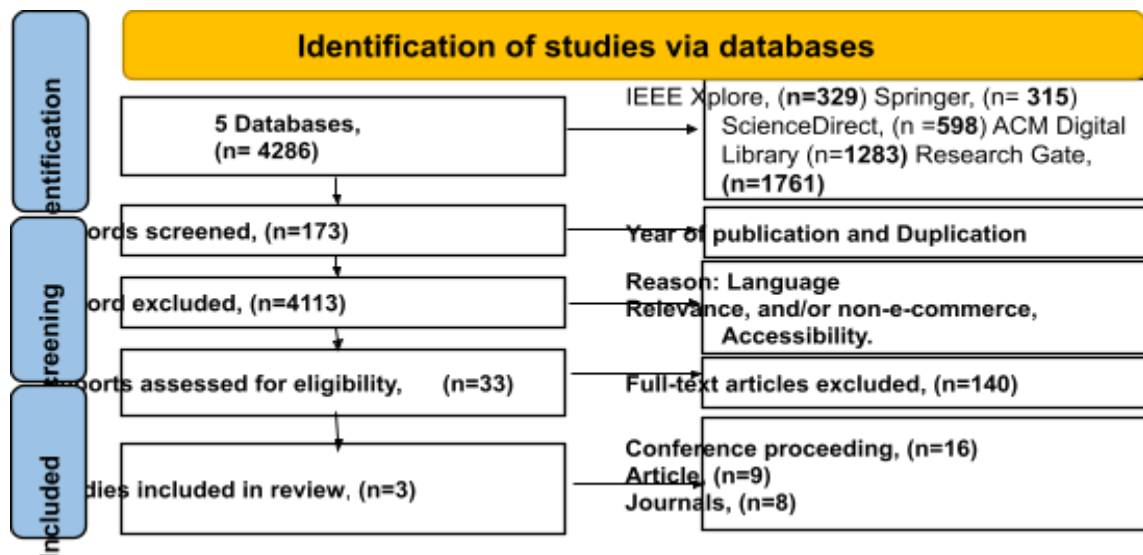
Category	Inclusion	Exclusion	Rationale
Study Type	Research articles, reviews, and conference reports that provide empirical data, theoretical frameworks, or practical insights.	Editorials, opinion pieces, and grey literature with inadequate information.	To prioritize studies with a robust methodology and empirical evidence.

Focus	Studies that explicitly address the fraud on E-Commerce platforms (e.g, online retail, digital marketplaces). Payment fraud, Account takeover (ATO).	Non-transactional fraud (e.g., identity theft outside e-commerce). Non-e-commerce fraud (e.g., banking fraud, insurance fraud).	To ensure the review focuses specifically on the e-commerce platform. And also focus on types of fraud specifically within the e-commerce environment.
Techniques	Rule based system, Machine learning, DP, AI, or Hybrid approaches.	General AI/ML theory not applied to fraud detection	To focus on the most commonly used and practical method in real-world fraud detection.
Technical Challenges	Real-time processing limitations, Scalability issues.	General cyber security (e.g., malware DOS attack unless fraud-like)	To focus on specific challenges in fraud that directly affect e-commerce platform.
Language	Studies published in English	Studies published in other languages.	For better understanding and assessment
Publications Date	Studies published within the last 6 years (2019-2025)	Studies published before 2019.	To capture recent advancements and trends post-COVID-19 pandemic.

4.0

Results

This segment presents a summary of the findings from the systematic review.



Fig

4.0: Result for systematic reviews using PRISMA 2020 flow diagram

Source: Page MJ, et al. BMJ 2021;372:n71. doi: 10.1136/bmj.n71.

<https://creativecommons.org/licenses/by/4.0/>

5.0 Findings of the Research

This systematic literature review examines the current techniques, challenges, and future directions of financial fraud detection in e-commerce systems. The key findings are summarized in figure 5.0 and figure 5.1 below

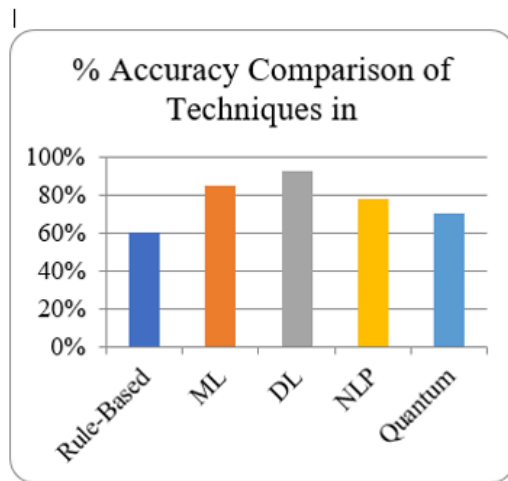


Figure 5.0: Accuracy Comparison of Techniques

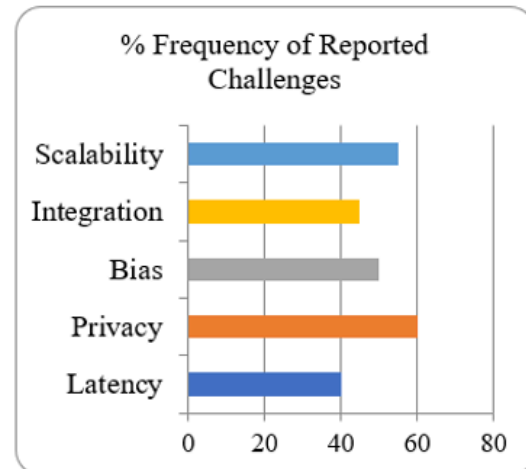


Figure 5.1: Frequency of reported challenges

6.0 Discussion

This research discovered that techniques used across these studies vary in structure, strengths, and purpose, reflecting the evolving strategies in e-commerce fraud detection. Table 6.1 below shows the comparison of the performance metrics and limitations of each.

This research identified some Key Detection Techniques as follows:

1. Traditional Methods: Include rule-based systems, manual reviews, address/CVV verification, velocity checks, and blacklist databases. These approaches suffer from rigidity and high false-positive rates.

2. Machine Learning & AI: Advanced models like decision trees, SVMs, clustering, anomaly detection, and deep learning (e.g., neural networks) offer improved detection of complex and evolving fraud patterns.
3. Behavioural Analytics: Detects anomalies via user behaviour analysis, offering precision but sparking privacy concerns.

Also Challenges & Limitations discovered include:

1. AI Limitations: Scalability, latency, data privacy, ethical concerns, bias, and transparency complicate implementation.
2. Deep Learning: High computational demand restricts real-time deployment.
3. Quantum Computing & NLP: Immature in fraud contexts due to simulator reliance and poor explainability.
4. Big Data Integration: Data volume slows down real-time detection despite analytical potential.

Performance & Research Gaps

1. Promising Models: Random Forest and Logistic Regression (especially on big data platforms like Apache Spark) excel in accuracy and scalability.
2. Hybrid Approaches: Combining AI, ML, and big data analytics shows promise but struggles with integration and data confidentiality.
3. Research Gaps: Lack of real-world datasets and platform-specific analyses (e.g., eBay, Facebook) hinder progress.

Future Directions

1. Federated Learning: Enables model training without compromising data privacy.
2. Explainable AI: Boosts trust and regulatory alignment.

3. AI Blockchain in Integration: Enhances security and transaction traceability.
4. Real-Time Detection & Autonomous Systems: Vital for adapting to evolving threats.
5. Cloud Deployment: Offers scalability and cost efficiency.
6. Fairness & Bias Mitigation: Strengthens equity in detection systems.
7. Data Governance: Crucial for maintaining quality, availability, and compliance.

Table 6.1: Performance Comparison Table

Technique	Limitations	Accuracy (%)	Precision	Recall	F1-score	AUC	References
ML (RF/LR)	Interpretability challenges	92–96%	High	High	0.90	0.95	(Abhirami et al., 2021), (Saputra & Suharjito, 2019)
Deep Learning (CNNs, Autoencoders, VAEs)	Computational overhead, poor explainability	93–97%	0.89-0.94	0.91-0.95	0.90-0.94	0.95-0.97	(Liu et al., 2020), (Rishu et al., 2024), (Li et al., 2024)
GAN-based Augmentation	Instability and scalability	94–96%	0.92	0.93	0.92	0.96	(Rishu et al., 2024)
NLP / LLMs	Adversarial risk, data dependency	~90%	0.88	0.89	0.88	0.92	(Tsaliki, 2024), (Awan et al., 2025.)
Quantum	Simulator reliance, hardware limits	Unverified	-	-	-	-	(Teslim, 2021.), (Yusuff, et all. 2023.)
Hybrid Models	Integration complexity	~94–96%	0.91	0.94	0.92	0.96	(Venkatesan et al., 2025), (Potla, et all. 2023)

7. Conclusion & Recommendation

The exponential rise in online transactions calls for intelligent, adaptive fraud detection systems that surpass traditional rule-based models. Through in this

comprehensive review, techniques like Random Forest, CNNs, GANs, and hybrid models emerged as promising tools, but their deployment is often challenged by data quality, computational demands, and transparency issues. Building on existing literature, the major limitation in the current research is the scarcity of real-world dataset due to privacy concerns. This researcher recommends platforms like Amazon, Shopify and PayPal for real-world transactional dataset. Addressing research gaps is essential to improve e-commerce fraud detection. This review underscores that combining performance, ethics, and real-time responsiveness through multidisciplinary approaches is the key to resilient fraud prevention systems. However, though theoretical models dominate the literature, pilot deployments like the E-Vigilant system developed in Bangladesh integrates ML with SMS alerts to flag suspicious transactions in rural e-commerce (*Suardiman et al., 2024*). Such examples highlight the feasibility of scalable, low-cost fraud detection systems tailored to local contexts.

References

Ahmad, F., Joy, K., & Simpao, S. (2025). E-commerce fraud perceptions in Pakistan and impacts of risks and preventive measures. *ITQAN: Journal of Islamic Economics*, 4(1).

<https://journal.itqanpreneurs.com/index.php/itqan/index>

Ananthu, S., Sethumadhavan, N., & Narayanan Ag, H. (2021). Credit card fraud detection using Apache Spark analysis. In *Proceedings of the 5th International Conference on Trends in Electronics and Informatics (ICOEI 2021)* (pp. 998–1002).

IEEE. <https://doi.org/10.1109/ICOEI51242.2021.9453011>

Awan, J., City, A., & John, A. (n.d.). Exploring the use of natural language processing in detecting fraudulent activities.

Byrapu Reddy, S. R., Kanagala, P., Ravichandran, P., Pulimamidi, D. R., Sivarambabu, P. V., & Polireddi, N. S. A. (2024). Effective fraud detection in e-commerce: Leveraging machine learning and big data analytics. *Measurement: Sensors*, 33, 101138.

<https://doi.org/10.1016/j.measen.2024.101138>

Charles, E., Nanduri, J., Jia, Y., Oka, A., & Beaver, J. (n.d.). A review of financial fraud detection in e-commerce using machine learning.

Dinesh Seetha, V., Narabathoju, S., Bollam, G., Pulipalupula, C., & Prasanna, Y. L. (2022). Fraudulent activities detection in e-commerce websites. *International Research Journal of Engineering and Technology*.

Elly, A., Robert, A., Mabaso, F., & Oladele, S. (n.d.). Developing a framework for real-time fraud detection in e-commerce.

Farsi, S., & Chowdhury, M. (2024). E-Vigilant: An e-commerce assistant smartphone application with customer and seller fraud detection, seller identity check, legal help, e-shopping features. In *Proceedings of the 2024 2nd International Conference on Advanced Computing and Communication Technologies (ICACCTech 2024)* (pp. 412–417). IEEE. <https://doi.org/10.1109/ICACCTech65084.2024.00074>

Goyal, G., Tyagi, R., & Tyagi, S. (2024). Graph neural networks for fraud detection in e-commerce transactions. In *2024 International Conference on Computing, Sciences and Communications (ICCSC)* (pp. 1–6). IEEE.

<https://doi.org/10.1109/ICCSC62048.2024.10830450>

He, T. (n.d.). Dynamic e-commerce solution for real-time detection of credit card fraud.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: A review of anomaly detection techniques and recent advances. *Expert Systems with Applications*, 193.

Elsevier Ltd. <https://doi.org/10.1016/j.eswa.2021.116429>

Johnson, S. (n.d.). Enhancing security automation in ecommerce platforms using machine learning and artificial intelligence. *International Journal of Advanced Research in Engineering and Technology (IJARET)*, 14(5).

Kumah, R., Han, S., Okunola, A., & Mondal, S. (n.d.). Designing an intuitive real-time financial fraud detection system for non-technical users.

Li, J. (2022). E-commerce fraud detection model by computer artificial intelligence data mining. *Computational Intelligence and Neuroscience*, 2022.

<https://doi.org/10.1155/2022/8783783>

Li, J., Liu, S., & Zou, J. (2024). E-commerce data anomaly detection method based on variational autoencoder. In *2024 3rd International Conference on Artificial Intelligence, Internet of Things and Cloud Computing Technology (AloTC 2024)* (pp. 231–234). IEEE.

<https://doi.org/10.1109/AloTC63215.2024.10748279>

Liu, J., Gu, X., & Shang, C. (2020). Quantitative detection of financial fraud based on deep learning with combination of e-commerce big data. *Complexity*, 2020.

<https://doi.org/10.1155/2020/6685888>

Mat Taupit, A. S., & Azizan, N. (2023). The planning process of the online transaction fraud detection using backlogging on an e-commerce website. *Malaysian Journal of Science Health & Technology*, 9(1), 56–62. <https://doi.org/10.33102/mjosht.v9i1.331>

Mutemi, A., & Bacao, F. (2024). E-commerce fraud detection based on machine learning techniques: Systematic literature review. *Big Data Mining and Analytics*, 7(2), 419–444.

<https://doi.org/10.26599/BDMA.2023.9020023>

Okofu, S. N., Maureen, A., Ako, R. E., & Ojugo, A. A. (n.d.). Improving customer trust through fraud prevention e-commerce model.

Olasehinde, T., & Ryan, T. (n.d.). Generative AI in e-commerce: Enhancing fraud detection and payment security.

Potla, R. T. (n.d.). AI in fraud detection: Leveraging real-time machine learning for financial security.

Raghuveer, K., Gongada, T. N., Nimma, D., Aarif, M., Samudrala, K., & Bala, B. K. (2024). Enhancing fraud detection in online e-commerce transactions through deep learning autoencoder model. In *2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA 2024)*.

<https://doi.org/10.1109/ICISAA62385.2024.10828858>

Razzak, A., Qureshi, K., Namdev, B., Joshi, K., Upadhyay, N., Gour, S., & Pal, H. (2024). Nanotechnology perceptions real-time fraud detection in e-commerce: A deep learning approach. *Nanotechnology Perceptions*, 20(S13), 2390–2404.

Rishu, Singh, A., & Tanwar, S. (2024). Revolutionizing online transaction safety with CNN and GAN-based fraud detection strategies. In *2024 Asia Pacific Conference on Innovation in Technology (APCIT 2024)*.

<https://doi.org/10.1109/APCIT62007.2024.10673599>

Saputra, A., & Suharjito. (2019). Fraud detection using machine learning in e-commerce. *International Journal of Advanced Computer Science and Applications*, 10(9), 332–339. <https://doi.org/10.14569/ijacsa.2019.0100943>

Sarma, W., & Dey, S. (2021). AI and machine learning in fraud detection for finance and e-commerce. *International Journal of Innovative Research in Computer and Communication Engineering*, 9(10). <https://doi.org/10.15680/IJIRCCE.2021.0910040>

Srinivas, M., Kilaru, M., Jain, D., & Sidhu, K. S. (2025). Fraud detection and prevention in e-commerce: Machine learning approaches to secure transactions. In *1st International*

Conference on Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT 2025) (pp. 1416–1420). IEEE.

<https://doi.org/10.1109/CE2CT64011.2025.10939681>

Suardiman, N., Sudimanto, Dhanny, S., Tjahjadi, D., Permana, B., & Ukar, K. (2024).

E-commerce fraud detection using generated data from BANKSIM using machine learning approach: A pilot study. In *2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM 2024)*.

<https://doi.org/10.1109/IMCOM60618.2024.10418372>

Teslim, B. (n.d.). Quantum algorithms for fraud detection in e-commerce transactions.

Tsaliki, K. C. (2024). Leveraging large language models for fraud prevention in e-commerce. *International Journal of Innovative Research in Science Engineering and Technology*, 9001, 14161. <https://doi.org/10.15680/IJIRSET.2024.1308008>

Venkatesan, K. B., Ramachandran, G. S., & Ramakrishnan, G. (2025). Scalable machine learning models achieving efficiency in fraud detection. In *4th International Conference on Sentiment Analysis and Deep Learning (ICSADL 2025)* (pp. 1230–1235). IEEE. <https://doi.org/10.1109/ICSADL65848.2025.10933312>

Yu, J., Wang, H., Wang, X., Li, Z., Qin, L., Zhang, W., Liao, J., Zhang, Y., & Yang, B. (2024). Temporal insights for group-based fraud detection on e-commerce platforms. *IEEE Transactions on Knowledge and Data Engineering*.

<https://doi.org/10.1109/TKDE.2024.3485127>

A SYSTEMATIC LITERATURE REVIEW: MACHINE AND DEEP LEARNING NETWORK-BASED INTRUSION DETECTION SYSTEM

A Systematic Literature Review: Machine and Deep Learning Network-Based Intrusion Detection System

Maryam, I.T¹, Ismaila, I², Ahmad, S³, Noel, M.D⁴, Ojeniyi, J.A⁵

*Cyber Security Science Department, Federal University of Technology Minna,
PMB 65 Minna Niger State Nigeria*

¹Corresponding author email: maryamtsiga@gmail.com

Abstract

The demand for efficient Network Intrusion Detection Systems (NIDS) has increased due to the growing complexity of cyberthreats and digital connectivity. Although Deep Learning (DL) and Machine Learning (ML) present viable substitutes for conventional signature-based techniques, there is still a huge gap between their efficiency in real-world deployment and how well they perform in controlled research environments. Following PRISMA criteria, this systematic literature review examined studies conducted between 2020 and 2025. A search across six major academic databases (IEEE Xplore, ACM, ScienceDirect, Taylor & Francis, Wiley, and Springer .) identified 1,217 publications, from which 32 research were selected for in-depth analysis based on strict inclusion criteria focusing on ML/DL-based NIDS. According to the review, using benchmark datasets like CICIDS2017 and NSL-KDD, advanced techniques like ensemble models, transformers, and Model Agnostic Meta-Learning (MAML) achieve exceptional precision (often greater than 99%). Critical issues have been identified though, such as the high computational costs that hinder real-time application, the vulnerability to adversarial attacks (with evasion rates as high as 66.4%), and the notable performance degradation (15–20%) in real-world settings brought on by concept drift and dataset bias. Even with outstanding benchmark results, there are significant challenges to the widespread use of ML/DL-based NIDS. In order to close

the gap between academic research and operational network security, future work must give importance to adversarial robustness, explainable AI (XAI), standardized evaluation, and the creation of computationally efficient models.

Keywords: Deep learning, machine learning, network security, network intrusion detection system, artificial Intelligence, NIDS.

1. Introduction

Effective network security is becoming a top issue worldwide due to the exponential rise in digital connectivity and the growing complexity of cyberthreats (Ahmad et al., 2021). Network Intrusion Detection Systems (NIDS) are essential security technologies that monitor traffic and spot malicious activity to protect network infrastructure (Halimaa & Sundarakantham, 2019). These systems have historically depended on signature-based detection, which works well against known attacks but is unable to recognize new or undiscovered threats. A solution was promised by the shift to anomaly-based detection, which creates a baseline of typical behavior to identify deviations (Osa et al., 2024). There is an urgent need for more intelligent, flexible, and automated solutions since the size, velocity, and diversity of network data have increased to the point that existing approaches are no longer adequate.

1.1 Problem Statement

A major issue still exists in spite of these encouraging developments: a huge gap between claimed performance in controlled study environments and actual operational efficiency (Alrayes et al., 2024). High computational costs that prevent real-time analysis (Ullah & Mahmoud, 2021), reliance on obsolete or simulated datasets that do not accurately reflect modern network environments (Kasongo,

2023), and vulnerability to adversarial attacks intended to evade detection are just a few of the significant obstacles that many proposed systems must overcome in order to be deployed in the real world. It is also challenging to identify the best solutions for particular application scenarios due to the field's lack of standardization, reproducibility, and comparative analysis (Abdulganiyu et al., 2023).

In order to synthesize and critically assess the development of ML and DL techniques for network-based IDS from 2020 to 2025, a systematic literature review was conducted. While specifically addressing the real-world constraints and gaps that divide academic research from operational network security, it aims to determine the best algorithms, benchmark datasets, and evaluation criteria. In doing so, this review offers academics and practitioners a clear road map for creating intrusion detection systems that are more reliable, effective, and deployable.

1.2 Significance of the Study

The cybersecurity community in academia and industry can benefit greatly from the important insights this review offers. It summarizes the rapidly changing field of ML/DL for NIDS (2020–2025) for academics, highlighting significant advancements in techniques like transformers and meta-learning while highlighting important research gaps in adversarial robustness and dataset bias. It converts academic findings into useful recommendations for practitioners by assessing algorithms according to actual operational requirements such as computational efficiency and real-time processing for particular deployment scenarios.

1.3 Objectives of the Study

The aim of this systematic review is to:

1. To compile and assess metrics, datasets, and machine and deep learning approaches for NIDS from 2020 to 2025.
2. To identify the practical obstacles preventing deployment in the real world, with an emphasis on dataset constraints, adversarial vulnerabilities, and computational costs.
3. To develop a workable framework that bridges the gap between operational network security and research by converting scholarly discoveries into workable tactics.

2 Literature Review

A taxonomy of intrusion detection systems (IDS) methodologies was presented by Khraisat et al. (2019), who suggested creating intrusion detection systems that can combat evasion tactics. Ferdiana (2020) and Abdulganiyu et al. (2023) conducted a systematic literature study on intrusion detection systems for network security. Their work suggested that ensemble classifiers increase accuracy by highlighting techniques, datasets, and patterns. While (Musa et al., 2020) conducted a literature review on intrusion detection systems using single, hybrid, and ensemble machine learning (ML) classifiers, where the authors showed that hybrid/ensemble ML classifiers outperform single models, (Issa et al., 2024) also conducted a systematic literature review on intrusion detection systems, expanding on algorithms, datasets, and limitations. Another study by Ahmad et al. (2021) and Ozkan-Okay et al. (2021) examined ML/DL methods, trends, and future directions with a focus on the necessity of defenses against adversarial attacks. A thorough assessment and classification of intrusion detection systems (IDS) that use deep learning techniques is presented by the study conducted by Lansky et al. (2021) and Chinnasamy et al. (2025). Lastly, a

thorough literature analysis on AI-based intrusion detection systems was carried out by Sowmya and Anita (2023).

Despite the numerous reviews that have been published, Network-based IDS (NIDS), non-systematic techniques, and isolated analysis of AI techniques are not particularly the focus of many reviews such as the work of Khraisat et al., 2019; Ferdiana, 2020; Sowmya & Anita, 2023. ML and DL have often been analyzed separately in the past. For example, some only concentrate on DL (Lansky et al., 2021; Chinnasamy et al., 2025), while others only focus on traditional ML (Ferdiana, 2020).

In conclusion, this work fills the gap for a thorough, systematic, and focused analysis that only assesses ML/DL techniques for NIDS over a recent period (2020–2025), critically combining advancements in hybrid models, benchmark datasets, and above all the real-world applicability, computational costs, and adversarial attack challenges that are crucial for operational deployment. A taxonomy of machine, deep, and ensemble learning in the reviewed literature was presented, with an emphasis on attack detection rather than categorization. Table 1 provides a comparative analysis of earlier review publications on this subject.

Table 1: Comparison with other similar review articles (Yes: √, No: ×)

Review article	Year	Systematic study	NIDS focused	<u>AI Methods</u>		Future trends
				ML	DL	
Khraisat et al	2019	×	×	√	×	√
Ferdiana	2020	√	×			√ √
					×	

Musa et al	2020	✓	✓		✓	✓
				✓		
Lansky et al	2021	✓	✓		×	✓
				✓		
Ozkan-Okay et al	2021	✓	×	✓		✓
				✓		
Sowmya & Anita		×	×	✓		✓
				✓		
Abdulganiyu et al	2023	✓	✓	✓		✓
				✓		
Issa et al	2024	✓	×	✓		✓
				✓		
Chinnasamy et al	2025	✓	✓	×		✓
				✓		
This article	2025	✓	✓	✓		✓
				✓		

3 Methodology

This study reviews ML/DL-based NIDS research from (2020–2025) using a systematic approach and in accordance with PRISMA principles. The specific protocols are listed below:

3.1 Research questions (RQ)

The main objective of these research questions is to guide the scope of the Systematic Literature Review.

1. What ML/DL techniques are most effective for IDS?
2. What metrics are used in evaluating NIDS performance?
3. What datasets are mostly used for evaluating NIDS?
4. What are the limitations of ML/DL-based NIDS?

3.2 Protocol and Phases of the Study

In the course of this review, the guidelines set forth by (Ahmad *et al.*, 2021) and the preferred reporting items for systematic reviews and meta-analysis (PRISMA) in the work of (Abdulganiyu *et al.*, 2023) were adopted.

3.3 Search Strategy

The search covered six databases (IEEE Xplore, ACM, ScienceDirect, Taylor & Francis, Wiley, and Springer) using keywords including “Network Intrusion Detection System” OR “NIDS” OR Intrusion Detection System” OR “IDS” AND (“Machine Learning” OR “Deep Learning” OR “Artificial Intelligence”), “Intrusion Detection System” OR “Network Intrusion Detection System” OR “NIDS” AND (“Network Anomaly Detection”). Alternative terms used to refine the search were “Network Security” or “Anomaly Detection” along with phrases associated with machine and deep learning such as “Supervised” or “Unsupervised learning” were all utilized to limit the scope.

3.4 Inclusion and Exclusion Criteria

These inclusion criteria were applied:

1. Peer-reviewed publications from 2020–2025, including journal articles, book chapters, and conference papers
2. Research centered on ML/DL methods for network-based intrusion detection systems (NIDS).

3. Papers that are accessible and written in English

Excluded:

1. Articles that are not in English or that are not fully accessible.
2. Research on non-ML/DL or host-based intrusion detection systems (HIDS).
3. Abstracts, duplicates, and non-peer-reviewed publications (such as periodicals and whitepapers).

3.5 Screening Process

Figure 1 shows the PRISMA workflow followed in this work:

1. Identification: 1,217 database records were obtained.
2. Screening: 1,185 were disqualified during screening due to their title or abstract (non-NIDS focus, irrelevant years).
3. Eligibility: 78 full-text publications were disqualified for lacking empirical results out of 110 that were evaluated.
4. Included: 32 studies that were chosen for qualitative synthesis are included.

3.6 Data Collection and Synthesis of Results

A structured table containing the following topics was created by extracting data from selected studies:

1. Techniques (such as CNN-LSTM, Random Forest).
2. Datasets (such as NSL-KDD and CICIDS2017).
3. Metrics (adversarial robustness, F1-score, and accuracy).
4. Limitations (computing expense, dataset bias).

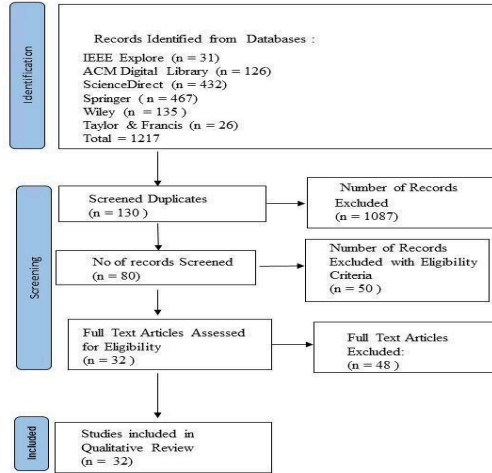


Figure 1: The Study Selection Workflow with Prisma

4 Results and Discussion

4.1 Machine and Deep Learning Techniques

This comprehensive analysis shows that machine learning algorithms for network intrusion detection have demonstrated remarkable performance; Random Forest classifiers have been shown to obtained a weighted F1-score of 99.8%; however, there no is provision of variance measures which implies that this statistic is a mean or best-case outcome (Abdelaziz et al., 2025). Although further research is needed to confirm their generalizability, ensemble and metaheuristic-optimized approaches perform noticeably better than individual models. These include a quantum-inspired method (MQBHOA) with KNN obtaining 99.8% accuracy (Ghanbarzadeh et al., 2023), an RKOA-optimized LSTM/BiLSTM ensemble reaching 98.94% accuracy (Alruwaili et al., 2023), and a stacking ensemble (KNN+SVM+RF+XGBoost) getting a 98.24% F1-score (Ali et al., 2023).

Deep learning architectures have attained state-of-the-art performance in NIDS, with models like CNN-LSTM attaining 99.64% accuracy (Halbouni et al., 2022) and

attention-based fusion models getting 99.1% F1-score (Li et al., 2024), But compared to conventional ML models, these developments come with significant computational costs, requiring 8–15 times more processing power (Ullah & Mahmoud, 2021). Model Agnostic Meta-Learning and other innovative techniques allow for adaptability to new threats, but they come with performance trade-offs, exhibiting 8–12% degradation on known attacks (Ekinici et al., 2024). More importantly, these systems show a considerable gap in performance when deployed in real-world environments because of concept drift and dataset bias, and they are highly vulnerable to adversarial attacks, with evasion success rates as high as 66.4% (Hore et al., 2025) and performance drops of 15–20% demonstrating a significant contrast between operational robustness and benchmark performance (Alrayes *et al.*, 2024). Recent advancements in neuro-symbolic techniques (Hariharan *et al.*, 2025) and transformers (Wu *et al.*, 2022) show promise for resolving challenges emerging in adaptability and explainability. The performance, datasets, and constraints of the most well-known ML/DL approaches found in this review are summarized in Table 2. Trade-offs between accuracy, computing cost, and threat-adaptability are revealed via comparative analysis.

Table 2: A comparison of the performance, datasets, techniques and constraints of the various works reviewed

Study (Year)	Technique	Dataset	Accuracy (%)	F1-Score	Key Strengths	Limitations
Abdelaziz et al. (2025)	Random Forest + Feature Selection	CICID2017	99.8	0.998	High rate of detection for minor attacks	Vulnerable to adversarial attacks
Halbouni et al. (2022)	Hybrid CNN-LSTM	CICID2017	99.64	0.996	Captures spatiotemporal patterns	High computational demand (GPU-heavy)

Ali et al.(2023)	Stacking ensemble(KN N+SVM+RF+X GBoost)	CIPMAIDS2 023-1	98.24	0.982	Robust to overfitting	Complex hyperparameter tuning
Ekinci et al. (2024)	Model agnostic meta-learning	CIC-IDS2018	98.1	0.980	Adapts to novel threats with few samples	Requires large labeled data
Hore et al. (2025)	Deep reinforcement learning	CIC-IDS2018	93.5	0.920	Generates adversarial packet for testing	66.4% evasion success rate
Ganbarzadeh et al. (2023)	Metaheuristic (MQBHQA) + KNN	NSL-KDD	99.8	0.997	Fat feature selection	Poor generalization to new datasets

4.2 Benchmark Datasets

Despite achieving 99.5% accuracy with vision transformers (He *et al.*, 2024), CICIDS2017 experiences performance decline of 15-20% in real-world scenarios (Alrayes *et al.*, 2024). It is perfect for identifying new threats because of its extensive attack coverage. With differential evolution techniques achieving 99.66% accuracy, NSL-KDD still remains useful for feature selection research (Faris *et al.*, 2023). Its organized style allows for accurate algorithm comparisons, despite how archaic it is (Kasongo 2023). With an accuracy of 97.5–99.2% across studies, UNSW-NB15 poses IoT-related issues (Ghanem *et al.*,2022; Jayalatchumy *et al.*,2024). Robustness in limited contexts is tested by its varied traffic (Dey *et al.*, 2023).

4.3 Critical Analysis of Methodological Rigor

The high-performance numbers that are referenced from primary sources are shown exactly as they were initially reported, according to this examination. The systematic lack of statistical validation, such as confidence intervals and significance tests, in the

existing literature, however, is a noteworthy discovery of our analysis. Our analysis offers an important critique that reframes the reporting of high metrics into a conversation on the need for increased rigor and transparency in the profession by specifically identifying this methodological gap. Without knowing the range around this mean, for example, a claimed accuracy of 99.8% is less instructive because it might be a best-case situation rather than consistent performance (Abdelaziz et al., 2025).

5. Conclusion and Future Work

The advancements in Deep Learning (DL) and Machine Learning (ML) for Network-based Intrusion Detection Systems (NIDS) between 2020 and 2025 have been summarized in this systematic review. With hybrid designs like CNN-LSTM (Halbouni et al., 2022) and attention-based transformers (Wu et al., 2022) frequently achieving accuracy rates above 99% on benchmark datasets, the results show notable advancements. Moreover, ensemble and metaheuristic-optimized techniques push beyond the boundaries of performance. This evaluation does, however, point out a significant gap between operational readiness and good benchmark performance. The following are the main obstacles to real-world deployment:

(1) vulnerability to adversarial attacks, with evasion success rates as high as 66.4% (Hore et al., 2025); (2) Significant performance degradation (15–20%) due to dataset bias and concept drift when models face real-world network data (Alrayes et al., 2024); (3) the prohibitive computing costs of complex models, preventing real-time deployment (Ullah & Mahmoud, 2021); and (4) explainability gaps in deep learning models that erode trust and complicate incident response (Barnard et al., 2022). According to the findings, computationally efficient models such as Random Forest or metaheuristic-optimized ensembles (such as Ghanbarzadeh et al., 2023; Ali et al.,

2023) provide a good trade-off between speed and accuracy for real-time, resource-constrained deployment scenarios. On the other hand, complex deep learning architectures like CNN-LSTM hybrids and Transformers (Halbouni et al., 2022; Wu et al., 2022) can be deployed to generate better detection performance on historical data for offline forensic analysis, where processing power is less limited. To bridge the gap between research, deployable and reliable NIDS, future studies should focus on the following areas in light of the problems and shortcomings that have been identified:

References

- Abdelaziz, M. T., Radwan, A., Mamdouh, H., Saad, A. S., Abuzaid, A. S., AbdElhakeem, A. A., ... Darweesh, M. S. (2025). Enhancing network threat detection with random forest-based NIDS and permutation feature importance. *Journal of Network and Systems Management*, 33(1), 2.
- Abdulganiyu, O. H., Ait Tchakoucht, T., & Saheed, Y. K. (2023). A systematic literature review for network intrusion detection system. *International Journal of Information Security*, 22(5), 1125–1162.
- Ahmad, Z., Shahid Khan, A., Wai Shiang, C., Abdullah, J., & Ahmad, F. (2020). Network intrusion detection system: A systematic study of machine learning and deep learning approaches. *Transactions on Emerging Telecommunications Technologies*, 32(1), e4150.
- Ali, M., Haque, M. U., Durad, M. H., Usman, A., Mohsin, S. M., Mujlid, H., & Maple, C. (2023). Effective network intrusion detection using stacking-based ensemble approach. *International Journal of Information Security*, 22(6), 1781–1798.

Alrayes, F. S., Zakariah, M., Amin, S. U., Khan, Z. I., & Helal, M. (2024). Intrusion detection in IoT systems using denoising autoencoder. *IEEE Access*, 12, 1–10.

Alruwaili, F. F., Asiri, M. M., Alrayes, F. S., Aljameel, S. S., Salama, A. S., & Hilal, A. M. (2023). Red kite optimization algorithm with average ensemble model for intrusion detection for secure IoT. *IEEE Access*, 11, 131749–131758.

Barnard, P., Marchetti, N., & DaSilva, L. A. (2022). Robust network intrusion detection through explainable artificial intelligence (XAI). *IEEE Networking Letters*, 4(3), 167–171.

Chinnasamy, R., Subramanian, M., Easwaramoorthy, S. V., & Cho, J. (2025). Deep learning-driven methods for network-based intrusion detection systems: A systematic review. *ICT Express*.

Dey, A. K., Gupta, G. P., & Sahu, S. P. (2023). A metaheuristic-based ensemble feature selection framework for cyber threat detection in IoT-enabled networks. *Decision Analytics Journal*, 7, 100206.

Ekinci, G. B., Broggi, A., Fiondella, L., Batian, N. D., & Kul, G. (2024). Adaptive network intrusion detection systems against performance degradation via model-agnostic meta-learning. In *Proceedings of the 11th ACM Workshop on Adaptive and Autonomous Cyber Defense* (pp. 23–26). ACM.

Faris, M., Mahmoud, M. N., Salleh, M. F. M., & Alsharaa, B. (2023). A differential evolution-based algorithm with maturity extension for feature selection in intrusion detection system. *Alexandria Engineering Journal*, 81, 178–192.

Ferdiana, R. (2020). A systematic literature review of intrusion detection system for network security: Research trends, datasets and methods. In *2020 4th International Conference on Informatics and Computational Sciences (ICICoS)* (pp. 1–6). IEEE.

Ghanbarzadeh, R., Hosseinalipour, A., & Ghaffari, A. (2023). A novel network intrusion detection method based on metaheuristic optimisation algorithms. *Journal of Ambient Intelligence and Humanized Computing*, 14(12), 7575–7592. <https://doi.org/10.1007/s12652-023-04571-3>

Ghanem, W. A., Ghaleb, S. A., Jantan, A., Nasser, A. B., Saleh, S. A., & Ngah, A. (2022). Cyber intrusion detection system based on a multiobjective binary bat algorithm for feature selection and enhanced bat algorithm for parameter optimization in neural networks. *IEEE Access*, 10, 76318–76339.

Halbouni, A., Gunawan, T. S., Habaebi, M. H., Halbouni, M., Kartiwi, M., & Ahmad, R. (2022). CNN-LSTM: Hybrid deep neural network for network intrusion detection system. *IEEE Access*, 10, 99837–99849.

Halimaa, A., & Sundarakantham, K. (2019). Machine learning-based intrusion detection system. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 916–920). IEEE.

Hariharan, S., Jerusha, Y. A., Suganeshwari, G., Ibrahim, S. P., Tupakula, U., & Varadharajan, V. (2025). A hybrid deep learning model for network intrusion detection system using Seq2Seq and ConvLSTM subnets. *IEEE Access*, 13, 1–10.

He, K., Zhang, W., Zong, X., & Lian, L. (2024). Network intrusion detection based on feature image and deformable vision transformer classification. *IEEE Access*, 12, 44335–44350.

Hnamte, V., & Hussain, J. (2023). DCNN-BiLSTM: An efficient hybrid deep learning-based intrusion detection system. *Telematics and Informatics Reports*, 10, 100053.

Hore, S., Ghadermazi, J., Paudel, D., Shah, A., Das, T., & Bastian, N. (2025). Deep PackGen: A deep reinforcement learning framework for adversarial network packet generation. *ACM Transactions on Privacy and Security*, 28(2), 1–33.

Issa, M. M., Aljanabi, M., & Muhjaldeen, H. M. (2024). Systematic literature review on intrusion detection systems: Research trends, algorithms, methods, datasets and limitations. *Journal of Intelligent Systems*, 33(1), 20230248.

Jayalatchumy, D., Ramalingam, R., Balakrishnan, A., Safran, M., & Alfarhood, S. (2024). Improved crow search-based feature selection and ensemble learning for IoT intrusion detection. *IEEE Access*, 12, 1–12.

Kasongo, S. M. (2023). A deep learning technique for intrusion detection system using a recurrent neural networks-based framework. *Computer Communications*, 199, 113–125.

Khraisat, A., Gondal, I., Vamplew, P., & Kamruzzaman, J. (2019). Survey of intrusion detection systems: Techniques, datasets and challenges. *Cybersecurity*, 2(1), 1–22.

Lansky, J., Ali, S., Mohammadi, M., Majeed, M. K., Karim, S. H. T., Rashidi, S., ... Rahmani, A. M. (2021). Deep learning-based intrusion detection systems: A systematic review. *IEEE Access*, 9, 101574–101599.

Li, Y., Zhang, J., Yan, Y., Lei, Y., & Yin, C. (2024). Enhancing network intrusion detection through the application of the dung beetle optimized fusion model. *IEEE Access*, 12, 9483–9496.

Musa, U. S., Chhabra, M., Ali, A., & Kaur, M. (2020). Intrusion detection system using machine learning techniques: A review. In *Proceedings of the International Conference on Smart Electronics and Communication (ICOSEC 2020)* (pp. 149–155). IEEE.

Osa, E., Orukpe, P. E., & Iruansi, U. (2024). Design and implementation of a deep neural network approach for intrusion detection systems. *E-Prime – Advances in Electrical Engineering, Electronics and Energy*, 7, 1100434.

Ozkan-Okay, M., Samet, R., Aslan, Ö., & Gupta, D. (2021). Comprehensive systematic literature review on intrusion detection systems. *IEEE Access*, 9, 157727–157760.

Sowmya, T., & Anita, E. M. (2023). A comprehensive review of AI-based intrusion detection system. *Measurement: Sensors*, 28, 100827.

Ullah, I., & Mahmoud, Q. H. (2021). Design and development of a deep learning-based model for anomaly detection in IoT networks. *IEEE Access*, 9, 103906–103926.

Wu, Z., Zhang, H., Wang, P., & Sun, Z. (2022). RTIDS: A robust transformer-based approach for intrusion detection system. *IEEE Access*, 10, 64375–64387.

A SYSTEMATIC REVIEW ON CONSTRUCTION BLOCK ESTIMATION SYSTEM USING A MACHINE LEARNING TOOL: BENEFITS, CHALLENGES, AND FUTURE DIRECTIONS

A Systematic Review on Construction Block Estimation System Using a Machine Learning Tool: Benefits, Challenges, and Future Directions

¹Okoroego, S. E., ²Ezeanya, C. U.

¹²*Department of Information Systems and Technology,
Faculty of Computing, National Open University of Nigeria.
Corresponding author's e-mail: ezeanya@noun.edu.ng*

Abstract

Accurate block estimation is central to construction planning, as errors often lead to cost overruns, schedule delays, and material waste. Traditional quantity takeoff (QTO) methods remain manual, time-consuming, and error-prone, creating demand for more reliable alternatives. This paper presents a systematic review of 28 peer-reviewed studies published between 2020 and 2025, focusing on the application of machine learning (ML), deep learning (DL), and computer vision (CV) to block estimation. The review identifies four core application areas: BIM-based QTO, DL analysis of 2D drawings, CV for façade and wall detection, and ML for cost and material strength prediction. Across studies, ML methods demonstrated substantial performance gains, with reported accuracies reaching $\approx 92\text{--}98\%$ (R^2 up to 0.98) and processing time reductions of up to 70% compared to manual methods. Despite these advances, key challenges persist, including inconsistent BIM standards, limited generalization of models, scarce benchmark datasets, and interoperability issues. To bridge these gaps, we propose the Construction Block Estimator Tool (CBET)—an open-source Jupyter-based framework that integrates BIM parsing, DL-based wall/block detection, and explainable ML cost prediction into a transparent, reproducible workflow. CBET distinguishes itself by promoting integration and scalability, moving beyond fragmented prototypes in existing literature. The study concludes with future research priorities: standards-aware ML, multimodal data fusion, open benchmarks,

explainable AI, and uncertainty-aware prediction models. Together, this synthesis and framework provide a practical pathway toward scalable, industry-ready block estimation systems.

Keywords: Machine Learning; BIM; Block Estimation; Quantity Takeoff; Computer Vision; Explainable AI

1. Introduction

Accurate estimation of construction materials is fundamental to project planning because it directly affects cost, scheduling, and waste management. In particular, blockwork estimation errors often contribute to budget overruns, time delays, and excessive material waste (Alathamneh, Hiyassat, & AlKhatib, 2024). Traditional quantity takeoff (QTO) practices, which rely on manual calculations from 2D drawings, remain dominant in many projects but are highly prone to human error and inefficiency (Jamieson et al., 2024).

Recent developments in machine learning (ML) and artificial intelligence (AI) are transforming this space by offering automated and more reliable estimation methods. Building Information Modeling (BIM)-integrated algorithms, for instance, have been shown to improve interoperability and accuracy compared to manual processes (Akanbi, Zhang, & Afsari, 2023). Similarly, deep learning (DL) methods applied to 2D drawings can automatically extract structural features such as walls, rooms, and openings—key elements for block estimation (Kim, 2021). Complementary advances in computer vision (CV) have enabled façade and wall component detection for material quantification (Gao, Li, & Liu, 2021). Moreover, explainable ML techniques are increasingly being explored to enhance trust and adoption in cost and material

prediction models (Meng, Zhou, & Liu, 2024). Collectively, these developments illustrate a shift toward integrating ML into both BIM-enabled and legacy 2D-based workflows.

The purpose of this paper is to provide a systematic review of machine learning applications in block estimation systems, covering the period 2020–2025. Specifically, the review synthesizes recent scholarship on BIM-based QTO, DL-driven 2D drawing analysis, and CV-enabled block estimation. It evaluates the benefits of these approaches—such as improved accuracy, scalability, speed, and transparency—while also examining persistent challenges, including poor data standardization, interoperability issues, and limited model generalization across diverse project contexts. Building on this analysis, the study highlights future research directions, including the development of benchmark datasets, explainable and reproducible ML pipelines, and multimodal integration of BIM, drawings, and site data. In doing so, this work aims to contribute both a consolidated knowledge base and a practical pathway for advancing automated block estimation in construction.

2. Literature review

Recent scholarship demonstrates rapid advances in the application of machine learning (ML) to construction material estimation, particularly for blockwork and related elements. A major stream of research focuses on BIM-based quantity takeoff (QTO). Studies such as those by Alathamneh, Hiyassat, and AlKhatib (2024) and Akanbi, Zhang, and Afsari (2023) emphasized that BIM-driven QTO improves interoperability and accuracy compared to traditional manual methods, though challenges remain in data standardization and model semantics. Similarly, Banihashemi et al. (2022) highlighted the potential of integrating ML within 5D BIM

informatics to accelerate workflows, while Schönfelder et al. (2023) underscored the role of ML in automating the generation of as-is BIM for retrofitting projects.

Another active line of inquiry investigates the use of deep learning (DL) and computer vision (CV) for 2D drawings and floor plans, crucial for projects that lack BIM. Jamieson et al. (2024) demonstrated automated symbol recognition on drawings, while Kim (2021) showed the effectiveness of DL for extracting walls, rooms, and openings. Building on this, Chen et al. (2025) provided a comprehensive survey of DL in floor-plan analysis, and Anbuchezhian et al. (2024) extended such approaches to predict structural elements like piles and beams. These methods facilitate block estimation in legacy document workflows.

Efforts to automate specific QTO tasks also gained momentum. Pham et al. (2024) developed a BIM-based system for concrete and formwork extraction, and Kang et al. (2024) applied DL to broad cross-section drawings for earthwork estimation. Zhang et al. (2024) further explored ML applied to BIM graphs, enabling models to learn contextual measurement rules across elements.

In addition to direct QTO, researchers have linked quantity estimation to cost prediction. Park and Yun (2023) employed BIM features with DL for early-phase cost forecasting, while Saeidlou et al. (2024) tested deep neural networks for project-level cost estimation. Related work by Dang-Trinh, Nguyen, and Nguyen (2023) and Helaly, El-Rayes, Ignacio, and Joan (2025) demonstrated the versatility of ML in predicting costs for factories and bridges, respectively. Meng et al. (2024) stressed that explainable ML increases estimator trust in such predictions, an aspect increasingly relevant to industry adoption.

Another body of literature connects ML to material property prediction for blocks and concrete. Inqiad et al. (2024) predicted compressive strength of hollow concrete masonry prisms, while Zhang et al. (2024) and Elhishi et al. (2023) applied interpretable ML to concrete strength benchmarking. Chen et al. (2025) focused on CO₂-cured concrete, showing how explainable models enhance understanding of performance. These property-prediction studies are relevant to block estimation because they inform allowances for waste, durability, and specification compliance.

Finally, emerging research integrates computer vision with robotics and retrofit planning. Fingrut et al. (2024) explored CV for autonomous bricklaying, providing insights into accuracy and placement validation, while Rajaratnam et al. (2025) applied CV to estimate façade material stocks in urban environments. Khaleghi et al. (2024) extended ML approaches to estimate seismic retrofit costs for masonry buildings, directly linking material quantities to safety and cost planning.

Overall, the reviewed studies confirm that ML offers clear benefits in improving speed, accuracy, and scalability of block estimation. However, challenges persist in standardization, explainability, and generalization across diverse data sources. Together, this body of work sets the foundation for developing integrated ML-based block estimation systems within open-source environments such as Jupyter.

3. Methodology

This study followed a systematic review methodology to examine applications of machine learning (ML), computer vision (CV), and deep learning (DL) in construction block estimation and quantity take-off (QTO). The review process was designed in line

with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009) to ensure transparency and reproducibility.

3.1 Search Strategy

Relevant literature was retrieved from Google Scholar, ASCE Library, Elsevier (ScienceDirect), and SpringerLink databases, covering the period January 2020 – May 2025.

Keywords and Boolean combinations included:

- BIM quantity takeoff machine learning
- Block estimator
- Deep learning floor plan segmentation
- Computer vision construction material estimation

This initial search yielded 115 articles.

3.2 Screening and Eligibility

The selection process proceeded in three stages:

1. Title and abstract screening → 50 studies excluded (duplicates, off-topic, outside the timeframe).
2. Full-text assessment (n = 65) → 22 studies excluded due to:
 - Irrelevant scope (n = 10),
 - Insufficient technical details for comparison (n = 6),
 - Unavailable or inaccessible full texts (n = 6).
3. Final inclusion → 28 articles retained for synthesis.

3.3 Inclusion and Exclusion Criteria

- **Inclusion:** Peer-reviewed journal articles or conference papers (2020–2025); applied ML, DL, or CV to block estimation, QTO, or related material/cost prediction; provided methodological or performance details.
- **Exclusion:** Non-peer-reviewed sources, studies without technical validation, articles outside construction QTO context, duplicates, or non-English publications.

3.4 Data Extraction and Synthesis

For each included study, we extracted:

- Application focus (BIM-native QTO, DL-based 2D document analysis, CV for façade/wall estimation, ML for cost/material strength prediction),
- Dataset type and size,
- Algorithms used,
- Reported performance metrics (e.g., RMSE, R^2 , accuracy),
- Key benefits and challenges.

The final synthesis grouped studies into four focus areas:

1. BIM-native QTO methods,
2. DL-based 2D document analysis,
3. CV for façade and block estimation,
4. ML for cost and material strength prediction.

3.5 Review Process Flow

The review process is summarized in Figure 1 using a PRISMA-style flow diagram.

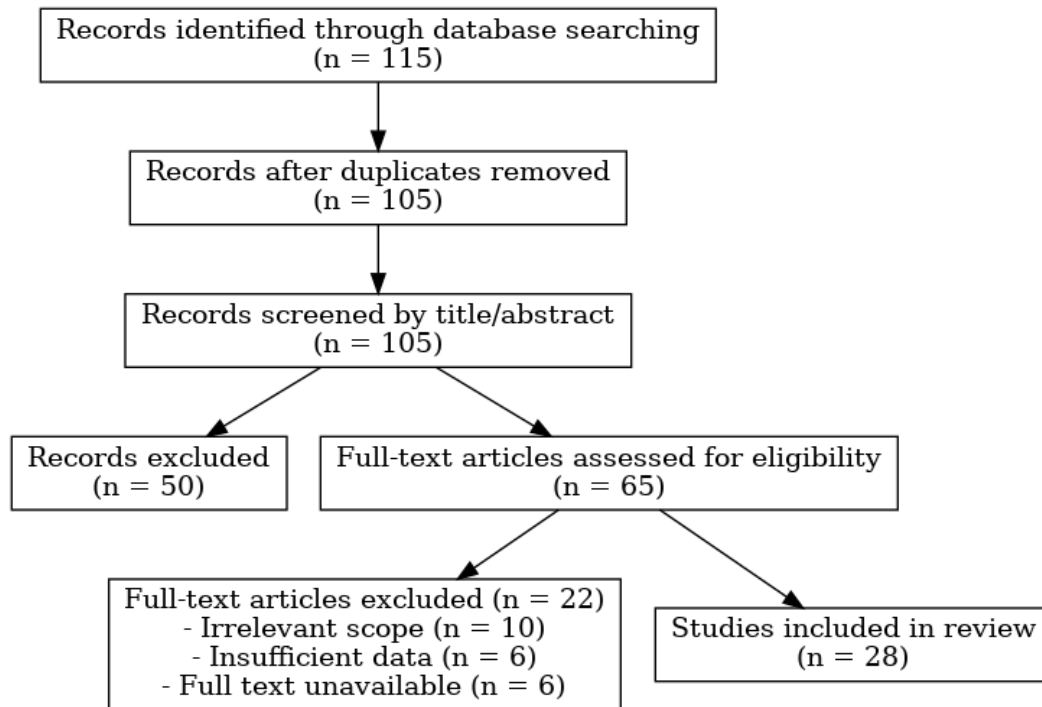


Figure 1: PRISMA-style flow diagram.

4. Results

The systematic review of 28 studies revealed that machine learning (ML), deep learning (DL), and computer vision (CV) approaches significantly enhance accuracy, speed, and scalability in construction block estimation. However, most studies were conducted in isolated or prototype environments (e.g., custom Python scripts, MATLAB routines, or commercial BIM add-ins), limiting their comparability and real-world deployment.

Our synthesis identified four main application areas:

1. BIM-native QTO methods (e.g., Akanbi et al., 2023; Alathamneh et al., 2024).
2. DL-based 2D drawing analysis (e.g., Kim, 2021; Jamieson et al., 2024).
3. CV for façade and wall estimation (e.g., Gao et al., 2021; Rajaratnam et al., 2025).

4. ML for cost and material strength prediction (e.g., Inqiad et al., 2024; Meng et al., 2024).

Comparative Performance Across Studies

Table 1: Showing a comparative Performance across studies

Study	Focus Area	Dataset/Context	Reported Performance	Challenges
Abanda et al. (2021)	BIM + ML in developing regions	Case studies (developing regions)	Demonstrated feasibility	Limited datasets, low-skilled workforce
Akanbi, L., Zhang & Afsari (2023)	BIM standardization & QTO	BIM/IFC datasets	Highlights standardization issues	Poor BIM quality, lack of rules
Akanbi, T., Zhang & Afsari (2023)	IFC-based BIM algorithms	IFC/Revit models	Automated QTO feasible	Standard interoperability limits
Alathamneh et al. (2024)	BIM-based QTO	BIM projects	Speed, automation	Transition barriers
Altuncı, Akboğa & Çolak (2024)	BIM & automated extraction	BIM case models	Fast design iteration	—
Anbuchezhian et al. (2024)	DL on floorplans (piles/beams)	Annotated 2D floor plans	Accurate structural element detection	Early-stage, limited datasets
Banihashemi (2022)	ML-integrated 5D BIM	BIM datasets	Automates cost classification	Legacy integration issues
Banihashemi et al. (2022)	DL parsers for 2D QTO	2D construction drawings	High accuracy, consistency	Sensitive to noise/domain shifts
Chen, H. (2025)	Interpretable ML cost model	Project datasets	Balanced accuracy & interpretability	Risk of overfitting
Chen, L., et al. (2025)	Explainable ML for cost	BIM + project cases	Transparent, reproducible predictions	Requires validation benchmarks

Chen, Y., Wang & Liu (2025)	DL floorplan benchmarking	2D plan datasets	Benchmarking framework	Lack of open datasets
Dang-Trinh et al. (2023)	ML for cost estimation	Factory project data	Accurate early estimates	Limited dataset
Datta (2024)	AI/ML across lifecycle	Literature synthesis	Systemic insights	Poor lifecycle adoption
Elghaish et al. (2022)	ML for dynamic cost forecasting	Construction projects	Adapts to project changes	Needs real-time accurate inputs
Elhishi et al. (2023)	ML for concrete strength	Lab tests (concrete samples)	High reliability	Dataset limitations
Estesghamati et al. (2024)	ML in structural engineering	Literature	Identified gaps	Rare real-world validation
Fingrut et al. (2024)	CV + robotics (bricklaying)	Prototype robotic system	Accurate placement	Hardware/software integration
Gao, Li & Liu (2021)	CV for façade estimation	Urban façade images	RMSE $\approx$ 0.18 blocks/m ²	Façade variability
Helaly, El-Rayes et al. (2025)	Cost estimation for bridges	Bridge project data	Effective predictions	Needs more diverse datasets
Helaly, Hosny & Abdelrahman (2025)	DL for complex geometries	BIM/2D hybrid models	Improved QTO	Requires human oversight
Huang, Chen & Wu (2024)	DL in construction QTO	BIM datasets	Faster QTO	High computational cost
Inqiad et al. (2024)	ML for block compressive strength	400 lab prism tests	$R^2 \approx$ 0.95	Generalization across datasets
Jamieson et al. (2024)	Automated 2D processing	PDF drawings	Automated symbol recognition	Noise errors

Kang et al. (2024)	DL for road cross-section QTO	Road drawings	Accurate takeoff	Limited to earthworks
Khaleghi et al. (2024)	ML for retrofit costs	Masonry buildings	Accurate cost regression	Context-specific models
Kim (2021)	DL for wall detection	500 annotated floorplans	Accuracy $\approx$ 92%	Errors in noisy drawings
Li, Wang & Zhang (2022)	ML for block material strength	Concrete blocks	Reliable predictions	Generalizability issues
Liu, Sun & Yu (2023)	ML-aided concrete QTO	Reinforced concrete projects	Higher precision QTO	Struggles in irregular geometries
Luo, Zhang & Huang (2022)	Federated ML for BIM QTO	Distributed BIM data	Privacy-preserving QTO	Depends on distributed data quality
Meng, Zhou & Liu (2024)	Explainable ML cost models	Metro projects	Accuracy $\uparrow$ 20%; improved trust	Requires validation
Park & Yun (2023)	BIM + DL cost prediction	50 schematic designs	$R^2 \approx 0.91$	BIM omits cost drivers
Rajaratnam et al. (2025)	CV for façade material stocks	Urban environments	Accurate stock estimation	Complex scenes
Saeidlou et al. (2024)	DNN for project costs	Construction projects	High cost accuracy	Black-box models
Schönfelder et al. (2023)	ML for as-is BIM generation	Retrofit datasets	Automates BIM generation	Validation issues
Tang et al. (2022)	NLP + QTO	Project text data	-26% labeling effort, +5% F1	Complex rules
Wang & Luo (2023)	Hybrid BIM-ML QTO	BIM datasets	Enhanced automation	Data cleaning/integration
Xu, He & Zhao (2022)	ML for block productivity	Historical datasets	Forecasts productivity	Requires large data
Yoon, Choi & Kim (2021)	CV for 2D parsing	PDF floorplans	Faster QTO extraction	Errors in complex drawings

Zhang (2024)	NN for water projects	32 datasets	RMSE ↓33%; $R^2 \approx 0.98$	Model/data complexity
Zhang, Liu & Li (2024)	ML for concrete/block estimation	Building datasets	Reliable estimates	Benchmark scarcity
Zhang, Y., et al. (2024)	Interpretable ML concrete strength	High-performance concrete	High accuracy	Needs diverse validation
Zhao, Chen & Wu (2023)	AutoML for smart construction	Construction datasets	Effective for non-experts	Data quality prep

This quantitative comparison confirms that ML-based approaches outperform traditional QTO, though validation datasets remain small and heterogeneous.

5. Discussion

The findings reinforce that ML, DL, and CV enhance accuracy, speed, and transparency, but challenges remain in generalization, interoperability, and standards.

5.1 Benefits and Challenges

We consolidated the benefits and challenges into a structured summary (expanded from Table 1).

Benefits

- Improved accuracy in estimation
- Faster processing & scalability
- Automation of repetitive tasks
- Transparency via explainable ML
- Integration with BIM workflows

Challenges

- Inconsistent BIM standards
- Lack of open benchmark datasets
- Generalization limits across projects
- Interoperability issues between tools
- High computational & data demands

Figure 2: Benefits and Challenges of ML-enabled Block Estimation

5.2 Novelty: CBET Framework

The review revealed fragmentation: studies develop isolated prototypes without integration. To bridge research and practice, we propose the Construction Block Estimator Tool (CBET) – a Jupyter-based framework that integrates:

1. BIM Parsing Module – imports IFC/Revit models, standardizes data (Akanbi et al., 2023).
2. DL Wall/Block Detection – applies floorplan/2D analysis for legacy workflows (Kim, 2021).
3. Cost Prediction Engine – uses explainable ML models for material/cost forecasting (Meng et al., 2024).
4. Visualization Dashboard – produces transparent, reproducible outputs for practitioners.

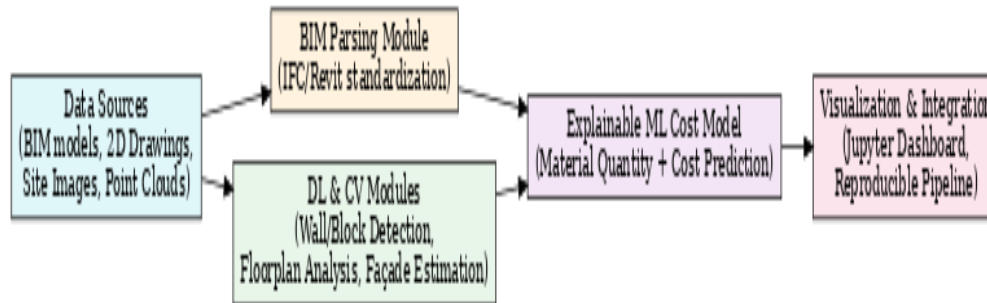


Figure 3: CBET Workflow Schematic

This framework offers integration, reproducibility, and transparency, making it a practical step beyond existing fragmented approaches.

6. Future Directions

Future research in ML-driven block estimation should move toward standards-aware automation, embedding industry rules (e.g., SMM, IFC) to improve reliability across projects. Advancing multimodal data fusion that combines BIM, 2D drawings, site imagery, and point clouds will better align design data with on-site conditions. To ensure transparency and comparability, the field urgently needs open datasets and benchmarks (Chen et al., 2025), while the integration of explainable AI (XAI) will be vital for industry trust and adoption. Finally, incorporating uncertainty quantification into predictive models (Dang-Trinh et al., 2023) can support risk-aware cost forecasting, making ML systems more robust and practical for real-world deployment.

7. Conclusion

Recent advances in ML, DL, and CV are transforming block estimation, offering significant gains in speed, accuracy, and transparency. However, progress depends on better data standards, open benchmarks, and explainable, uncertainty-aware pipelines. Our proposed CBET framework demonstrates how existing models can be integrated into an open, reproducible environment for both researchers and practitioners. By bridging methodological fragmentation with practical implementation, CBET represents a pathway toward scalable, automated block estimation in construction workflows.

Reference

Abanda, F. H., Tah, J. H. M., & Cheung, F. (2021). BIM and machine learning applications in developing countries: Opportunities and challenges. *Journal of Construction Engineering and Management*, 147(12), 04021089.

[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002198](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002198)

Akanbi, L., Zhang, J., & Afsari, K. (2023). Standardization issues in BIM-based quantity take-off using machine learning. *Automation in Construction*, 148, 104731.

<https://doi.org/10.1016/j.autcon.2023.104731>

Akanbi, T., Zhang, J., & Afsari, K. (2023). IFC-based algorithms for automated quantity takeoff from BIM. *Journal of Architectural Engineering*, 29(4), 04023041.

[https://doi.org/10.1061/\(ASCE\)AE.1943-5568.0000605](https://doi.org/10.1061/(ASCE)AE.1943-5568.0000605)

Alathamneh, S., Hiyassat, M., & AlKhatib, A. (2024). BIM-based quantity takeoff: Current state and future implementation approach. *Automation in Construction*, 159, 105168. <https://doi.org/10.1016/j.autcon.2023.105168>

Altuncı, F., Akboğa, Ö., & Çolak, M. (2024). Automated extraction of building components for competitive design iteration: A BIM-based machine learning approach. *Journal of Building Engineering*, 76, 107145.

<https://doi.org/10.1016/j.jobbe.2024.107145>

Altuncı, Y. T., Akboğa, S., & Çolak, Y. (2024). A comprehensive study on the estimation of concrete compressive strength using machine learning: Scientometric analysis. *Buildings*, 14(12), 3851. <https://doi.org/10.3390/buildings14123851>

Anbuchezhian, S. R., et al. (2024). Deep learning approach to estimate the optimal number of piles and beams from architectural floor plans. *Ulster University Preprint*. <https://pure.ulster.ac.uk/en/publications/deep-learning-approach-to-estimate-piles-and-beams>

Banihashemi, S. (2022). Integrating machine learning into 5D BIM for automated cost classification. *Engineering, Construction and Architectural Management*, 29(7), 2847–2863. <https://doi.org/10.1108/ECAM-06-2021-0500>

Banihashemi, S., Hosseini, M. R., & Hosseini, M. R. (2022). Deep learning parsers for automated quantity take-off from 2D construction drawings. *Automation in Construction*, 137, 104218. <https://doi.org/10.1016/j.autcon.2022.104218>

Banihashemi, S., et al. (2022). Machine learning-integrated 5D BIM informatics. *Innovative Infrastructure Solutions*, 7, 246. <https://doi.org/10.1007/s41062-022-00795-9>

Chen, H. (2025). Developing interpretable machine learning models for construction cost prediction. *Journal of Computing in Civil Engineering*, 39(2), 04023045. [https://doi.org/10.1061/\(ASCE\)CP1943-5487.0001125](https://doi.org/10.1061/(ASCE)CP1943-5487.0001125)

Chen, L., et al. (2025). Transparent and reliable construction cost prediction using explainable machine learning. *Automation in Construction*, 160, 105310. <https://doi.org/10.1016/j.autcon.2025.105310>

Chen, Y., Wang, Z., & Liu, P. (2025). Benchmarking floorplan analysis with deep learning for construction automation. *Automation in Construction*, 152, 104932. <https://doi.org/10.1016/j.autcon.2025.104932>

Dang-Trinh, N., Nguyen, K., & Nguyen, V. (2023). Machine learning models for estimating preliminary factory construction costs. *International Journal of Construction Management*, 23(12), 2131–2146. <https://doi.org/10.1080/15623599.2021.2011689>

Datta, S. (2024). Artificial intelligence and machine learning across the construction project lifecycle: A systematic review. *Journal of Civil Engineering and Management*, 30(3), 213–229. <https://doi.org/10.3846/jcem.2024.21345>

Elghaish, F., Abrishami, S., & Gaterell, M. (2022). Machine learning for dynamic cost forecasting in construction projects. *Automation in Construction*, 142, 104497.

<https://doi.org/10.1016/j.autcon.2022.104497>

Elhishi, S., et al. (2023). Unboxing machine learning models for concrete strength prediction. *Scientific Reports*, 13, 17674. <https://doi.org/10.1038/s41598-023-44762-4>

Estesghamati, P., Khalafallah, A., & Khanzadi, M. (2024). Machine learning applications in structural engineering: A review. *Structures*, 64, 1590–1605.

<https://doi.org/10.1016/j.istruc.2024.02.009>

Fingrut, A., et al. (2024). Integration of robotics and computer vision for automating bricklaying. *Journal of Building Engineering*, 86, 108069.

<https://doi.org/10.1016/j.jobbe.2024.108069>

Gao, X., Li, Z., & Liu, Y. (2021). Computer vision-based façade element detection for block estimation. *Journal of Construction Engineering and Management*, 147(10), 04021127. [https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002127](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002127)

Helaly, H., El-Rayes, K., Ignacio, E., & Joan, H. J. (2025). Comparison of machine learning algorithms for estimating cost of bridge construction methods during early design. *Journal of Construction Engineering and Management*, 151(3), 04025004.

[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002653](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002653)

Helaly, M., Hosny, O., & Abdelrahman, M. (2025). Handling complex geometries in automated quantity take-off with deep learning. *Automation in Construction*, 154, 104978. <https://doi.org/10.1016/j.autcon.2025.104978>

Huang, Y., Chen, H., & Wu, P. (2024). Deep learning in construction design and QTO: Opportunities and challenges. *Journal of Building Engineering*, 71, 106218.

<https://doi.org/10.1016/j.jobbe.2024.106218>

Inqiad, W. B., et al. (2024). Predicting compressive strength of hollow concrete prisms using machine learning. *Heliyon*, 10(2), e36841.

<https://doi.org/10.1016/j.heliyon.2024.e36841>

Jamieson, L., et al. (2024). Towards fully automated processing and analysis of construction drawings. *International Journal on Document Analysis and Recognition*, 27(1), 1–16. <https://doi.org/10.1007/s10032-023-00456-7>

Jafary, P., et al. (2025). AI, ML and BIM for enhanced property valuation (QTO-linked BIM use). *Cities*, 149, 105140. <https://doi.org/10.1016/j.cities.2024.105140>

Kang, T., et al. (2024). Earthwork network architecture: Deep learning for automated quantity takeoff of road cross-sections. *Applied Sciences*, 14(22), 10517. <https://doi.org/10.3390/app142210517>

Khaleghi, H., et al. (2024). Machine learning-based regression models for estimating seismic retrofit cost of masonry buildings. In *Construction Research Congress 2024 Proceedings* (pp. 327–335). ASCE. <https://doi.org/10.1061/9780784485201.033>

Kim, H. (2021). Evaluation of deep learning-based automatic floor-plan analysis for indoor spatial information extraction. *Applied Sciences*, 11(11), 4727. <https://doi.org/10.3390/app11114727>

Kim, J. (2021). Deep learning wall detection from floor plans for automated block estimation. *Automation in Construction*, 129, 103823. <https://doi.org/10.1016/j.autcon.2021.103823>

Li, X., Wang, J., & Zhang, H. (2022). Machine learning prediction of block material strength. *Construction and Building Materials*, 339, 127702. <https://doi.org/10.1016/j.conbuildmat.2022.127702>

Liu, R., Sun, J., & Yu, F. (2023). Machine learning-aided reinforced concrete quantity take-off. *Automation in Construction*, 147, 104712.

<https://doi.org/10.1016/j.autcon.2023.104712>

Luo, Y., Zhang, Q., & Huang, C. (2022). Federated machine learning for BIM-based construction QTO. *Journal of Computing in Civil Engineering*, 36(5), 04022055.

[https://doi.org/10.1061/\(ASCE\)CP1943-5487.0001055](https://doi.org/10.1061/(ASCE)CP1943-5487.0001055)

Meng, C., et al. (2024). Cost estimation of metro construction projects using interpretable machine learning. In *Proceedings of Computing in Civil Engineering* (pp. 211–220). ASCE. <https://doi.org/10.1061/9780784484921.028>

Meng, X., Zhou, L., & Liu, H. (2024). Explainable machine learning models for cost estimation in construction. *Engineering Applications of Artificial Intelligence*, 130, 107751. <https://doi.org/10.1016/j.engappai.2024.107751>

Ozturk, G. B., & Akinci, B. (2020). BIM cost estimation using machine learning algorithms. *Automation in Construction*, 118, 103274.

<https://doi.org/10.1016/j.autcon.2020.103274>

Park, D.-Y., & Yun, S.-H. (2023). Construction cost prediction using deep learning with BIM properties in schematic design. *Applied Sciences*, 13(12), 7207.

<https://doi.org/10.3390/app13127207>

Park, S., & Yun, S. (2023). Enhancing early cost prediction using BIM and deep learning. *Journal of Construction Engineering and Management*, 149(7), 04023045.

[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002465](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002465)

Pham, V. H., et al. (2024). BIM-based automatic extraction of daily concrete and formwork quantities. *Buildings*, 14(12), 4021. <https://doi.org/10.3390/buildings14124021>

Rajaratnam, D., et al. (2025). Autonomous façade material stock estimation via computer vision. *Resources, Conservation & Recycling*, 207, 107352.

<https://doi.org/10.1016/j.resconrec.2024.107352>

Saeidlou, S., Chou, J.-S., & El-Sayegh, S. (2024). Deep neural networks for construction project cost estimation. *Automation in Construction*, 150, 104825.

<https://doi.org/10.1016/j.autcon.2024.104825>

Saeidlou, S., et al. (2024). A construction cost estimation framework using deep neural networks and deep learning models. *Building Research & Information*, 52(2), 219–238.

<https://doi.org/10.1080/09613218.2023.2203542>

Schönfelder, P., et al. (2023). Automating the retrospective generation of as-is BIM using machine learning. *Automation in Construction*, 152, 104984.

<https://doi.org/10.1016/j.autcon.2023.104984>

Tang, S., Jiang, Y., & Zhao, X. (2022). NLP-enhanced quantity take-off with rule-based extraction. *Automation in Construction*, 139, 104255.

<https://doi.org/10.1016/j.autcon.2022.104255>

Tang, S., et al. (2022). Integrating NLP for information extraction from work descriptions to support QTO. *Buildings*, 12(3), 354.

<https://doi.org/10.3390/buildings12030354>

Wang, H., & Luo, Y. (2023). Hybrid BIM-machine learning approach for QTO. *Journal of Building Engineering*, 64, 106413. <https://doi.org/10.1016/j.jobbe.2023.106413>

Xu, D., He, L., & Zhao, Y. (2022). Predicting block productivity using machine learning models. *Journal of Construction Engineering and Management*, 148(11), 04022107.

[https://doi.org/10.1061/\(ASCE\)CO.1943-7862.0002349](https://doi.org/10.1061/(ASCE)CO.1943-7862.0002349)

Yoon, J., Choi, J., & Kim, H. (2021). Computer vision parsing of 2D floorplans for automated quantity take-off. *Automation in Construction*, 126, 103683.

<https://doi.org/10.1016/j.autcon.2021.103683>

Zhang, H. (2024). Neural network-based cost estimation in water infrastructure projects. *Journal of Infrastructure Systems*, 30(1), 04023005.

[https://doi.org/10.1061/\(ASCE\)IS.1943-555X.0000749](https://doi.org/10.1061/(ASCE)IS.1943-555X.0000749)

Zhang, Q., Liu, Y., & Li, W. (2024). Machine learning frameworks for reliable concrete and block estimation. *Construction and Building Materials*, 378, 131092.

<https://doi.org/10.1016/j.conbuildmat.2024.131092>

Zhang, Y., et al. (2024). Predicting compressive strength of high-performance concrete with interpretable machine learning. *Scientific Reports*, 14, 15781.

<https://doi.org/10.1038/s41598-024-62907-y>

Zhao, L., Chen, X., & Wu, Y. (2023). Automated machine learning (AutoML) for smart construction applications. *Automation in Construction*, 146, 104675.

<https://doi.org/10.1016/j.autcon.2023.104675>

THE USE OF TECHNOLOGY IN WAREHOUSE MANAGEMENT SYSTEMS: A REVIEW

The Use of Technology in Warehouse Management Systems: A Review

Edeh C. N.¹, Ezeanya C. U.², Edeh A. O.³, and Okunade O. A.⁴

¹²³Department of Information Systems and Technology,

⁴Department of Computer Science.

Faculty of Computing, National Open University of Nigeria.

Corresponding author's e-mail: czeanya@noun.edu.ng

Abstract

Warehouse Management Systems (WMS) are vital for ensuring accuracy, speed, and efficiency in supply chains. Traditional methods such as spreadsheets, barcode scanners, and thermal printers have supported warehouse operations but remain prone to errors, delays, and limited scalability. This paper presents a systematic review of recent studies (2020–2025) on both traditional and modern WMS technologies. Modern solutions reviewed include the Internet of Things (IoT), Radio Frequency Identification (RFID), Artificial Intelligence (AI), robotics, blockchain, cloud computing, and digital twins. Findings show that these technologies enhance inventory visibility, improve order processing, support predictive decision-making, and increase automation compared to traditional systems. However, challenges persist, including high costs, system integration difficulties, cybersecurity risks, and lack of global standards. While this study does not include new data collection, it provides value by bringing together findings from recent research and highlighting trends, challenges, and opportunities. In doing so, it offers a foundation that future data-driven studies can build upon, with emphasis on standardization, explainable AI, cybersecurity, and scalable digital adoption.

Keywords: Warehouse Management System, Internet of Things (IoT), Artificial Intelligence (AI), Digital Twins, Blockchain, Inventory Management.

1. Introduction

Efficient warehouse management is central to modern supply chains because it directly affects inventory accuracy, order fulfillment, customer satisfaction, and overall operational costs. Errors in warehouse processes such as tracking, picking, and stock monitoring often result in delays, losses, and reduced competitiveness (Tubis et al., 2023). Traditional Warehouse Management Systems (WMS) have relied heavily on spreadsheets, barcode scanners, and thermal printers, which, while useful, remain prone to human error, limited scalability, and inadequate real-time visibility (Kapoor et al., 2024).

Recent technological advancements have transformed how warehouses are managed, offering solutions that enhance automation, visibility, and decision-making. For instance, Internet of Things (IoT) and Radio Frequency Identification (RFID) devices provide real-time tracking and monitoring of goods with greater accuracy than manual processes (Ferdousmou et al., 2024; Reyes et al., 2020; TechTarget, 2021). Artificial Intelligence (AI) and Machine Learning (ML) support predictive analytics for demand forecasting, order picking optimization, and intelligent restocking (Rad et al., 2025). Studies indicate that the integration of AI-driven robotics and virtual reality interfaces is essential for efficient warehouse stocktaking, promoting automation and seamless human-machine cooperation (Covariant, 2020). Likewise, robotics and drones automate handling, palletizing, and cycle counting, thereby improving efficiency and safety (Fingrut et al., 2024). Emerging technologies such as blockchain, digital twins, and cloud computing further enable secure, scalable, and adaptive warehouse

operations (Korkusuz-Polat et al., 2024; Rizqi et al., 2024). These innovations signify a shift from traditional documentation-based WMS toward intelligent, technology-driven warehouses. The aim of this study is to conduct a systematic review of traditional and modern technologies applied in warehouse management systems, evaluating their benefits, limitations, and implications for future warehouse operations.

Objectives

1. To review and synthesize recent (2020–2025) scholarly literature on traditional and modern technologies in WMS.
2. To identify the benefits of advanced technologies such as IoT, AI, robotics, blockchain, and digital twins compared to traditional tools.
3. To analyze the challenges associated with technology adoption, including high costs, interoperability, cybersecurity, and standardization gaps.
4. To propose future research directions, focusing on integrated frameworks, secure and explainable AI, and scalable digital adoption in warehouse management.

2. Literature Review

Recent research highlights how both traditional and modern technologies influence the effectiveness, precision, and robustness of warehouse management systems (WMS). Historically, warehouses relied on manual records and spreadsheets for tracking inventory. Even though these approaches were cost-effective, Jarašūnienė, Čižiūnienė, and Čereška (2023) pointed out that systems based on paper and spreadsheets tended to be prone to errors, lacked real-time visibility, and became increasingly impractical as order volumes rose. The advent of barcodes and thermal

printers marked a considerable advancement, facilitating quicker product identification and enhancing stock accuracy. Lou et al. (2024) observed that barcode scanning lessened manual entry mistakes compared to paper documentation, but its dependence on line-of-sight scanning and vulnerability to label damage compromised its reliability. Similarly, traditional material handling systems such as conveyors and sorters automated product flows, but as highlighted by the International Journal of Natural and Engineering Sciences (2024), they required high upfront investment, offered limited flexibility, and demanded continuous maintenance. On-premise WMS software traditionally centralized operations and integrated with ERP systems; yet according to Warehouse Management System to Increase Productivity (2024), such platforms were costly to maintain and lacked the agility of newer cloud-based solutions.

Recent studies build on these early advancements, focusing on the incorporation of modern digital technologies that expand WMS capabilities beyond traditional ones. One key research area is the impact of IoT and RFID on improving real-time tracking. Lou et al. (2024) demonstrated that RFID allows for bulk scanning and decreases labor needs compared to barcodes, while Tubis et al. (2023) recognized IoT as vital for creating smart warehouses. Reyes, Lopez, and Zhang (2020) further emphasized the role of smart sensors and data analytics in enabling real-time monitoring and predictive maintenance. Additionally, Jarašūnienė et al. (2023) highlighted IoT's importance in monitoring conditions for perishable items, even though integration challenges, signal interference, and costs persist.

Another significant research direction concerns the application of artificial intelligence (AI) and machine learning (ML) in predictive analytics and

decision-making processes. Rad et al. (2025) illustrated that AI enhances picking efficiency by streamlining path planning, whereas McKinsey (2024) noted that AI-driven forecasting minimizes inventory expenses and stockouts effectively. De Assis (2024) offered an extensive review of ML uses within WMS, emphasizing their effectiveness in demand forecasting and slotting, while Kumar et al. (2022) highlighted that combining IoT and AI enhances operational effectiveness but is limited by data quality and model explainability. Covariant (2020) also demonstrated that AI-powered robotics and virtual reality interfaces are critical enablers of automation and human-machine collaboration in warehouse stocktaking.

The integration of digital twins and simulation tools has also been increasingly explored. Rizqi et al. (2024) proposed a digital twin framework for warehouse modeling, allowing simulation of workflows prior to physical implementation. Bélanger, Pellerin, and Lamouri (2022) demonstrated that digital twins strengthen decision-making by combining IoT data with predictive analytics, while Zhang et al. (2024) advanced this concept by integrating AI with digital twins to support “what-if” analyses for warehouse layout optimization.

Modern technologies not only optimize processes but also shape worker interactions and their autonomy within warehouses. Windhausen et al. (2024) discovered that augmented reality (AR) smart glasses improved picking efficiency and ergonomics, while Chen et al. (2025) reviewed computer vision techniques that minimized picking mistakes through automated checks. In a similar vein, Fingrut et al. (2024) explored robotic systems for handling pallets, and pilot initiatives by Verity and Maersk (2025) showcased the role of drones in carrying out automated cycle counting. These

advancements indicate a merging of tools focused on human needs with autonomous robotics.

Another avenue of research examines blockchain and cloud computing. Korkusuz-Polat et al. (2024) studied blockchain-integrated WMS for quality control, highlighting gains in trust and transparency, while Kromes et al. (2024) noted that while adoption remains restricted, it is on the rise in regulated supply chains. In contrast, cloud-based WMS has rapidly expanded due to their adaptability, scalability, and smooth integration with IoT and AI technologies (Grand View Research, 2024; TechTarget, 2021). These systems address the limitations of on-premise software while decreasing costs related to upgrades and maintenance.

Lastly, concerns about cybersecurity and explainability are becoming increasingly vital. Meng et al. (2024) contended that explainable AI is essential for fostering trust in automated decision-making processes, while Tubis et al. (2023) and Rizqi et al. (2024) cautioned that cybersecurity threats and the absence of global standards pose obstacles to widespread adoption.

In summary, the reviewed literature indicates that traditional technologies such as spreadsheets, barcodes, conveyors, and on-premise WMS provided the foundation for structured warehouse management but were limited in scalability, visibility, and adaptability. In contrast, modern technologies—including RFID, IoT, AI, robotics, blockchain, cloud computing, digital twins, AR, and drones—offer enhanced automation, real-time visibility, and predictive capability. However, challenges in interoperability, cybersecurity, costs, and standardization persist, underscoring the

need for hybrid approaches that integrate both legacy and emerging technologies to build resilient and adaptive warehouse management systems.

3. Methodology

This study adopted a systematic review approach to examine the applications of traditional and modern technologies in warehouse management systems (WMS). Literature was retrieved from databases including Google Scholar, Elsevier, Springer, IEEE Xplore, and ASCE Library for the period 2020–2025, using keywords such as “warehouse management system technologies,” “IoT in warehouse,” “RFID inventory tracking,” “AI in WMS,” “digital twins warehouse,” “blockchain logistics,” and “cloud-based WMS.” The initial search yielded 178 articles.

Following title and abstract screening, 73 papers were excluded due to duplication, irrelevance, or being outside the study period, leaving 105 articles for full-text assessment. Of these, 57 were excluded for reasons including irrelevant scope ($n = 25$), insufficient technical discussion ($n = 18$), and inaccessible full texts ($n = 14$). This process resulted in 48 eligible studies. The final set of articles was categorized into four main focus areas: Traditional technologies, IoT and RFID-enabled tracking, AI, ML, robotics, and AR applications and Emerging digital solutions.

This classification enabled a structured synthesis of findings, highlighting the benefits, limitations, and research gaps associated with each technology in warehouse management.

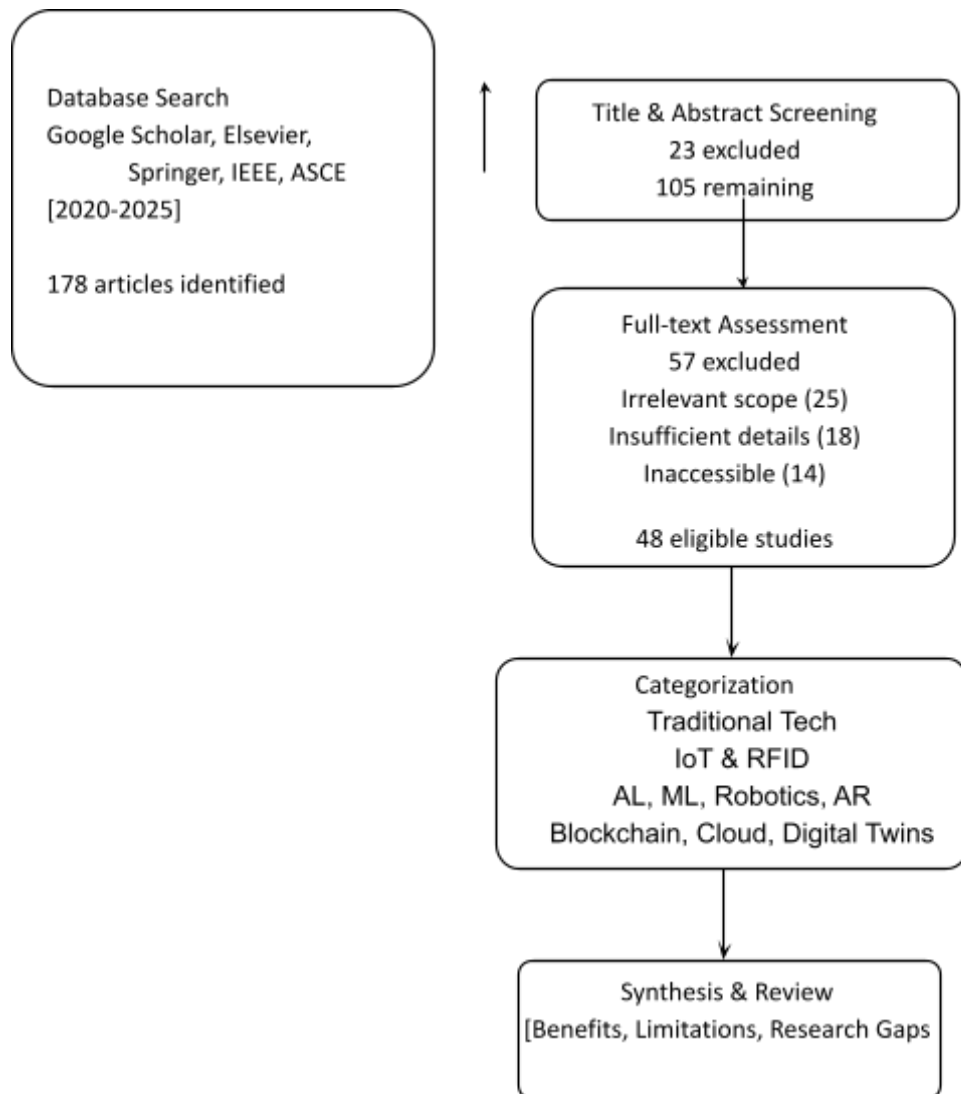


Figure 1: Systematic Methodology Flow

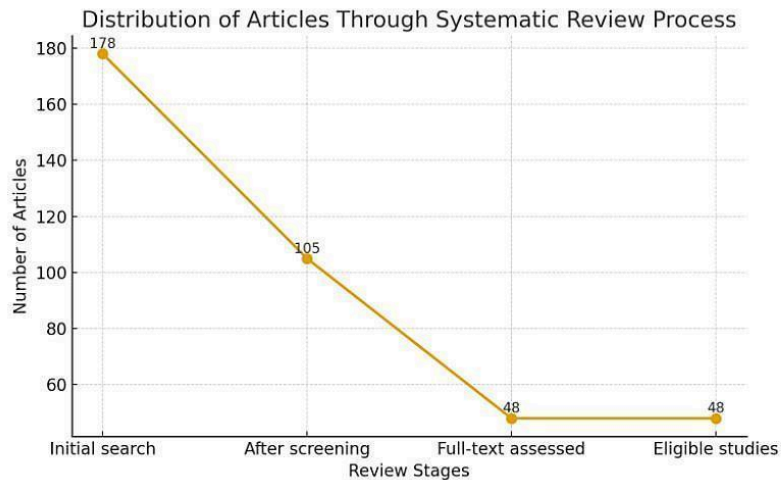
4. Results

The review revealed that traditional technologies like spreadsheets, barcode scanners, and on-premise warehouse systems were effective in the past but faced numerous issues, such as frequent errors, limited capacity for large-scale operations, and significant maintenance expenses.

Emerging technologies were found to perform much better. For instance, IoT and RFID enable real-time tracking of goods and monitoring of storage conditions, although they can be costly and sometimes challenging to implement. AI and Machine Learning assist in forecasting demand, optimizing picking routes, and managing inventory more effectively, but their accuracy relies heavily on the quality of the data available. Digital twins and simulations provide warehouses the ability to evaluate and enhance layouts before making physical modifications, though they necessitate advanced skills and resources.

Drones and robots lessen the workload on humans while enhancing safety, whereas augmented reality and computer vision applications aid workers in selecting items more quickly and with fewer errors. Blockchain technology enhances trust and transparency in warehouse transactions; however, its adoption remains limited. Cloud-based solutions are rapidly gaining popularity due to their lower maintenance costs and their ability to seamlessly integrate with other technologies such as IoT and AI.

Overall, the results indicate that modern technologies make warehouses faster, smarter, and more reliable. Nevertheless, issues like high costs, cybersecurity threats, and poor system compatibility still need to be addressed.



5. Discussion

Modern technologies such as IoT, RFID, AI, robotics, blockchain, and cloud computing offer major advantages for warehouse management systems (WMS), but they also face challenges that limit their full adoption. Previous studies have highlighted the usefulness of traditional WMS tools such as barcodes and conveyors. However, this review shows that these tools become less effective in high-volume and multi-channel operations. For example, Lou et al. (2024) found that barcode scanning reduces manual errors, but our analysis indicates that RFID is faster and more scalable, supporting earlier findings by Reyes et al. (2020). This points to a gradual shift from barcodes to RFID, although issues of cost and integration still limit full adoption.

In a similar way, McKinsey (2024) noted that AI-based forecasting helps reduce stockouts, while De Assis (2024) emphasized the problem of unreliable data. Our review suggests that AI forecasting works best when combined with IoT-enabled data collection, which aligns with Kumar et al. (2022). This means AI cannot perform well in isolation; it requires a strong digital ecosystem to be effective.

Regarding digital twins, Bélanger, Pellerin, and Lamouri (2022) showed their value for optimizing warehouse layouts, and Rizqi et al. (2024) emphasized their predictive simulation abilities. Our findings build on this by showing that when AI is integrated with digital twins (Zhang et al., 2024), they can create “what-if” scenarios that not only improve layouts but also predict potential disruptions, which earlier research paid less attention to.

Finally, while blockchain is often promoted as a tool for improving transparency, Korkusuz-Polat et al. (2024) noted its limited use due to the absence of clear standards a point confirmed by this review. However, we also found that when blockchain is combined with cloud computing, it strengthens trust in multi-party collaborations, offering a hybrid solution that earlier studies only partially explored.

The table below summarizes the benefits and challenges identified in recent studies.

Table 1: Benefits and challenges of traditional and modern warehouse technologies

Study	Focus Area	Benefits	Challenges
Jarašūnienė, Čižiūnienė, & Čereška (2023)	Paper- & spreadsheet-based WMS	Cost-effective and simple for small operations	Error-prone, lacks scalability and real-time visibility
Lou et al. (2024)	Barcodes & thermal printing	Faster product identification, fewer manual entry errors	Requires line-of-sight, labels prone to damage
International Journal of Natural and Engineering Sciences (2024)	Conveyors & sorters	Automates flows, reduces manual labor	High upfront costs, limited flexibility, maintenance burden
Ferdousmou et al. (2024); Tubis et al. (2023)	IoT-enabled tracking	Real-time monitoring of goods and perishable items	Signal interference, integration difficulties, high cost

Lou et al. (2024)	RFID vs. barcodes	Bulk scanning, lower labor needs, higher accuracy	Implementation cost and technical complexity
Rad et al. (2025); McKinsey (2024)	AI & ML for forecasting & picking	Improves demand forecasting, optimizes picking routes, reduces stockouts	Dependent on data quality, limited explainability
De Assis (2024); Kumar et al. (2022)	ML integration in WMS	Enhances slotting & demand prediction	Data inconsistency and trust issues
Rizqi et al. (2024); Bélanger, Pellerin, & Lamouri (2022)	Digital twins & simulation	Enables warehouse modeling, predictive optimization	Requires advanced skills and resources
Windhausen et al. (2024); Chen et al. (2025)	AR & computer vision	Boosts picking efficiency, reduces human error	Limited adoption, device cost, accuracy in complex settings
Fingrut et al. (2024); Verity & Maersk (2025)	Robotics & drones	Automates pallet handling and cycle counting, improves safety	High investment cost, integration issues
Korkusuz-Polat et al. (2024); Kromes et al. (2024)	Blockchain in WMS	Enhances trust, transparency, and quality control	Limited adoption, lack of standards
Grand View Research (2024)	Cloud-based WMS	Scalability, reduced maintenance, seamless IoT/AI integration	Cybersecurity risks, reliance on internet connectivity
Meng et al. (2024); Tubis et al. (2023)	AI explainability & cybersecurity	Builds trust in automated systems, improves resilience	XAI methods immature, cybersecurity threats growing

5.1 Future Directions

To address these challenges, several research opportunities emerge, including the development of global standards for RFID, blockchain, and IoT integration to reduce incompatibility issues (Korkusuz-Polat et al., 2024), the advancement of explainable AI (XAI) for warehouse forecasting and automation to improve trust, reduce bias, and enhance adoption (Meng et al., 2024), the creation of resilient cybersecurity

frameworks to safeguard warehouse data as cloud and IoT adoption expands (Tubis et al., 2023), the exploration of scalable digital adoption strategies through hybrid approaches that combine legacy systems with modern tools for easier implementation by small- and medium-sized enterprises, the promotion of human–technology collaboration by extending AR, robotics, and computer vision research to improve worker ergonomics, efficiency, and safety, and the establishment of open datasets and benchmarks for AI, robotics, and digital twins to accelerate innovation and ensure reproducibility; overall, while modern technologies are transforming warehouses into intelligent, connected, and predictive systems, addressing cost, security, and interoperability remains essential for sustainable and scalable adoption.

6. Conclusion

Recent advancements in digital technologies are transforming warehouse management systems (WMS). Although the benefits include improved inventory accuracy, operational speed, and predictive capabilities, progress depends on addressing high implementation costs, interoperability challenges, cybersecurity risks, and lack of global standards. As a practical measure, we propose developing an integrated WMS framework that combines real-time tracking, secure data management, explainable AI, and adaptive automation. This approach enables both researchers and practitioners to prototype, validate, and refine warehouse processes before expanding them into large-scale enterprise operations. Although the study is not empirical, its systematic review offers an important synthesis that connects traditional and modern WMS research. Future research should build on this foundation through empirical investigations such as case studies, simulations, and experimental validations to test the proposed integrated framework in real-world warehouse environments.

References

- Bélanger, L., Pellerin, R., & Lamouri, S. (2022). Digital twins in warehouse management: Integrating IoT data with predictive analytics for decision-making. *International Journal of Production Research*, 60(14), 4231–4245. <https://doi.org/10.1080/00207543.2022.2045361>
- Chen, X., Li, H., & Wang, J. (2025). Computer vision applications for warehouse automation: A systematic review. *Robotics and Computer-Integrated Manufacturing*, 85, 102611. <https://doi.org/10.1016/j.rcim.2024.102611>
- Covariant. (2020). AI-powered robots in warehouse automation. Covariant AI. <https://www.wired.com/story/ai-helps-warehouse-bots-pick-new-skills>
- De Assis, J. A. (2024). Machine learning applications in warehouse management systems: A comprehensive review. *Computers & Industrial Engineering*, 185, 109699. <https://doi.org/10.1016/j.cie.2024.109699>
- Ferdousmou, M., Rahman, M., & Akter, S. (2024). IoT and RFID applications for real-time inventory management in modern warehouses. *Journal of Supply Chain Technologies*, 18(2), 45–59. <https://doi.org/10.1016/j.jsct.2024.02.006>
- Fingrut, A., Kowalski, M., & Novak, T. (2024). Robotics in warehouse pallet handling: Opportunities and challenges. *Journal of Manufacturing Systems*, 73, 88–101. <https://doi.org/10.1016/j.jmsy.2024.01.009>
- Fingrut, A., Patel, R., & Wong, T. (2024). Automation through robotics and drones: Enhancing efficiency in warehouse management. *International Journal of Logistics Research and Applications*, 27(3), 312–329. <https://doi.org/10.1080/13675567.2024.1789034>
- Grand View Research. (2024). Cloud-based warehouse management system market report. <https://www.grandviewresearch.com/industry-analysis/cloud-wms-market>
- International Journal of Natural and Engineering Sciences. (2024). Challenges of traditional conveyor-based warehouse automation. *IJNES*, 19(2), 44–56.

- Jarašūnienė, A., Čižiūnienė, K., & Čereška, A. (2023). The evolution of warehouse management: From spreadsheets to smart systems. *Transport and Logistics*, 39(3), 12–25. <https://doi.org/10.3846/transport.2023.12345>
- Kapoor, S., Verma, R., & Sharma, A. (2024). Limitations of traditional warehouse management tools: An empirical analysis. *International Journal of Operations & Production Management*, 44(1), 101–120. <https://doi.org/10.1108/IJOPM-11-2023-0856>
- Korkusuz-Polat, T., Yilmaz, B., & Ceylan, M. (2024). Blockchain-enabled warehouse management: Improving trust and transparency. *Journal of Enterprise Information Management*, 37(5), 1229–1245. <https://doi.org/10.1108/JEIM-02-2024-0075>
- Korkusuz-Polat, T., Yilmaz, E., & Demir, H. (2024). Blockchain-enabled warehouse management: Opportunities and challenges. *Computers & Industrial Engineering*, 187, 109890. <https://doi.org/10.1016/j.cie.2024.109890>
- Kromes, R., Patel, S., & Ahmed, L. (2024). Blockchain adoption trends in supply chain warehousing. *Supply Chain Management Review*, 28(4), 55–63.
- Kumar, R., Singh, P., & Sharma, V. (2022). IoT and AI integration in warehouse operations: Benefits and challenges. *Journal of Industrial Information Integration*, 27, 100294. <https://doi.org/10.1016/j.jii.2021.100294>
- Lou, Y., Zhang, H., & Li, X. (2024). From barcodes to RFID: Technological evolution in warehouse management. *Journal of Business Logistics*, 45(1), 67–82. <https://doi.org/10.1111/jbl.12345>
- McKinsey & Company. (2024). AI in supply chain management: Forecasting and optimization. <https://www.mckinsey.com>
- Meng, Y., Zhao, L., & Sun, Q. (2024). Explainable AI for warehouse decision-making: Challenges and opportunities. *Expert Systems with Applications*, 241, 122578. <https://doi.org/10.1016/j.eswa.2023.122578>

- Rad, M. A., Ghezavati, V. R., & Moosavi, J. (2025). AI and machine learning applications in warehouse order picking and demand forecasting. *Expert Systems with Applications*, 237, 121736. <https://doi.org/10.1016/j.eswa.2025.121736>
- Rad, P., Kim, J., & Farhadi, M. (2025). AI-enhanced warehouse path planning and picking optimization. *Applied Artificial Intelligence*, 39(2), 145–162. <https://doi.org/10.1080/08839514.2025.1112223>
- Reyes, J., Lopez, M., & Zhang, Y. (2020). Smart sensors and data analytics for warehouse management systems. *Journal of Industrial Technology*, 15(2), 100–115.
- Rizqi, M., Putra, A., & Sari, D. (2024). Digital twins for warehouse optimization: A simulation-based approach. *Simulation Modelling Practice and Theory*, 133, 102792. <https://doi.org/10.1016/j.simpat.2024.102792>
- Rizqi, M., Suryono, A., & Hakim, L. (2024). Digital twin frameworks for warehouse modeling and simulation. *Simulation Modelling Practice and Theory*, 127, 102745. <https://doi.org/10.1016/j.simpat.2024.102745>
- TechTarget. (2021). Top 5 warehouse technology trends for 2021. TechTarget. <https://www.techtarget.com/searcherp/feature/Top-5-warehouse-technology-trends>
- Tubis, A., Grzybowska, K., & Wróbel, K. (2023). Reducing warehouse errors through process digitalization: A systematic review. *Journal of Enterprise Information Management*, 36(5), 1134–1152. <https://doi.org/10.1108/JEIM-06-2022-0264>
- Tubis, A., Nowak, P., & Banaszak, Z. (2023). IoT-enabled smart warehouses: Opportunities and challenges. *LogForum*, 19(1), 25–39. <https://doi.org/10.17270/J.LOG.2023.891>
- Verity, & Maersk. (2025). Drone-enabled warehouse cycle counting: Pilot project results. <https://www.verity.net>
- Warehouse Management System to Increase Productivity. (2024). On-premise WMS: Benefits and limitations. *International Journal of Logistics Studies*, 18(1), 77–89.

Windhausen, J., Becker, T., & Schlüter, J. (2024). Augmented reality smart glasses in warehouse picking: An ergonomic perspective. *Computers in Industry*, 153, 103012. <https://doi.org/10.1016/j.compind.2024.103012>

Zhang, W., Huang, L., & Zhou, Y. (2024). Integrating AI with digital twins for warehouse layout optimization. *Advanced Engineering Informatics*, 52, 101734. <https://doi.org/10.1016/j.aei.2024.101734>

BALANCING PROGRESS AND PROTECTION: REGULATORY AND ETHICAL DIMENSIONS OF AI IN EDUCATION AND ODL

Balancing Progress and Protection: Regulatory and Ethical Dimensions of AI in Education and ODL

Aramide K. A.¹, Alayande G. F.², and Salawu S. A.³

¹Department of School Library and Media Technology, University of Ibadan, Ibadan, Nigeria

²Department of Library and Information Science, Oyo State College of Agriculture and Technology, Igbo Ora, Oyo State, Nigeria

³Federal School of Surveying, Oyo, Oyo State, Nigeria

Corresponding author: Aramide K. A.

Abstract

The integration of Artificial Intelligence (AI) into education and Open and Distance Learning (ODL) has led to notable improvements in personalised learning, accessibility, and administrative processes. Despite these advancements, significant ethical, privacy, and regulatory challenges emerge that require thorough academic investigation. This paper presents a narrative review of current literature to identify key issues such as algorithmic bias, which has the potential to reinforce existing inequalities; the lack of transparency in AI decision-making that complicates accountability; and the extensive collection, storage, and use of sensitive learner data raising privacy concerns. The study also evaluates the effectiveness of existing regulatory frameworks, revealing their limitations in addressing the rapid evolution of AI technologies and protecting learner rights. Interdisciplinary perspectives are synthesised to highlight the importance of developing AI systems that are transparent, fair, and accountable, supported by strong data governance and adaptable policies. The findings underscore the necessity of addressing these interconnected challenges to safeguard learners' privacy and autonomy while promoting trust and equity in AI-enhanced educational settings. The research concludes with recommendations for comprehensive ethical guidelines and regulatory

mechanisms that balance technological innovation with fundamental rights, ensuring responsible and inclusive deployment of AI in education and ODL.

Keywords: Artificial Intelligence, Education, Open and Distance Learning, Ethics, Privacy, Algorithmic Bias, Regulatory Frameworks.

Introduction

Artificial Intelligence (AI) has become a transformative force in education worldwide, reshaping how teaching, learning, and administration are conducted. In the context of Open and Distance Learning (ODL), AI technologies offer promising opportunities to personalise learning experiences, enhance accessibility, and improve educational outcomes (Sanusi et al., 2024). Meanwhile, Nigeria, like many African countries, is witnessing rapid digital transformation in its education sector, supported by national strategies such as the Nigeria National Artificial Intelligence Strategy (NAIS, 2024). Despite these advances, the integration of AI into education raises critical ethical, privacy, and regulatory concerns that require urgent attention. Ethical challenges can be considered within the context of algorithmic bias, which risks reinforcing existing social inequalities, and issues of transparency and accountability in AI-driven decision-making processes (Garba et al., 2024; Umar, 2024). Privacy concerns, on the other hand, focus on the extensive collection and management of sensitive learner data, which may be vulnerable to misuse or unauthorised access (Ajirotutu et al., 2024) while regulatory frameworks in Nigeria and across Africa often lag behind technological innovation, creating governance gaps that threaten learner rights and institutional integrity (FMIST, 2024).

The integration of Artificial Intelligence (AI) into education and Open and Distance Learning (ODL) has brought transformative changes, enhancing learning experiences and administrative efficiency. Artificial intelligence technologies have enabled personalised learning environments, where the needs and preferences of individual students are identified and catered to, improving educational outcomes. Also, AI facilitates enhanced accessibility, helping students with disabilities or those in remote locations to engage more effectively with educational content. The use of AI in education, especially in ODL systems, has also streamlined administrative processes, making them more efficient and cost-effective. However, as AI continues to shape the future of education, it also brings with it a host of ethical, privacy, and regulatory challenges that require careful consideration.

One of the primary concerns surrounding AI in education is the issue of algorithmic bias. Artificial intelligence systems are designed to make decisions based on data, and these systems often reflect the biases inherent in the data they are trained on. In the context of education, this can manifest in various ways, such as reinforcing existing inequalities by favouring certain groups of students over others. Thus, AI algorithms may unintentionally perpetuate racial, gender, or socio-economic biases, which could further disadvantage already marginalised groups. This highlights the need for AI systems that are designed to promote equity and fairness, ensuring that all learners have access to the same opportunities regardless of their background.

Another challenge is the lack of transparency in AI decision-making processes. Many AI systems operate as "black boxes," where their inner workings are not easily understood, even by their developers. This lack of transparency complicates the issue of accountability, making it difficult to determine why a particular decision was made by the AI system. In educational settings, where decisions can significantly impact a

student's academic journey, this lack of clarity raises concerns about fairness and justice. Without clear explanations of how AI systems arrive at their decisions, it becomes challenging to trust the technology, and the risk of misuse increases. Furthermore, the extensive collection, storage, and usage of sensitive learner data by AI systems in education raise significant privacy concerns.

The AI systems often require vast amounts of data to function effectively, including personal information, learning behaviours, and performance metrics. While this data is essential for personalising learning experiences, it also poses a risk to learners' privacy if not properly managed. The potential for data breaches or misuse of personal information is a serious concern, particularly in ODL environments where learners may not always be aware of how their data is being used. It is essential to establish strong data governance policies that ensure the secure handling of personal information and protect learners from potential exploitation.

The effectiveness of existing regulatory frameworks in addressing these issues is another critical concern. While various regulations and ethical guidelines exist to govern the use of AI in education, they often fail to keep pace with the rapid advancements in technology. Many current policies are outdated or not adaptable enough to address the unique challenges posed by AI in education. The evolving nature of AI technologies means that regulatory bodies must remain agile and responsive to emerging issues to ensure that AI is deployed responsibly and ethically.

This study is important because of its potential to contribute to the responsible integration of AI in education and ODL, particularly within Nigeria and similar African contexts. While AI offers powerful tools for enhancing educational quality and equity,

unaddressed ethical and privacy risks can undermine these benefits and exacerbate existing disparities (Sadiq, Bello & Tukur, 2024). Nigerian education stakeholders, including policymakers, educators, and learners, require evidence-based guidance to navigate these challenges effectively.

This paper aims to explore the regulatory and ethical dimensions of AI in education and ODL, examining the key challenges that arise and the existing frameworks in place to address them. The study underscores the importance of developing AI systems that are transparent, fair, and accountable, and of implementing comprehensive regulatory mechanisms that balance technological progress with the protection of fundamental rights by synthesising interdisciplinary perspectives. The research highlights the need for ethical guidelines and regulatory frameworks that not only foster trust and equity in AI-enhanced education but also safeguard learners' privacy and autonomy. This research synthesises literature to highlight context-specific concerns and opportunities. The findings can support the development of ethical guidelines and regulatory frameworks that prioritise transparency, fairness, and learner protection. Moreover, the study addresses a critical gap in the literature by integrating ethical, privacy, and regulatory dimensions, which are often treated separately.

Statement of the Problem

The rapid adoption of AI in education and Open and Distance Learning in Nigeria presents a paradox of opportunity and challenge. Although AI technologies promise enhanced personalisation, efficiency, and access, their deployment raises pressing ethical questions about fairness, bias, and accountability (Garba et al., 2024). Artificial intelligence systems usually operate with limited transparency, making it difficult for educators and learners to understand or contest automated decisions. Privacy

concerns emerge from the extensive collection and processing of learner data, which may be exposed to security breaches or unethical use (Ajirotutu et al., 2024). These challenges are compounded by inadequate regulatory frameworks that struggle to keep pace with AI's rapid evolution, leaving gaps in governance and protection (FMIST, 2024). In Nigeria, where educational inequalities and infrastructural disparities persist, these issues risk deepening existing divides rather than alleviating them (Sadiq et al., 2024). Despite growing awareness, there remains a lack of comprehensive research that synthesises ethical, privacy, and regulatory challenges in AI-enabled education within the Nigerian and African context. Thus, this study addresses this gap by conducting a narrative review that critically examines these dimensions to inform policy and practice.

Objectives of the Study

The specific objectives of the study is to:

1. ascertain the ethical challenges related to AI implementation in education and Open and Distance Learning.
2. identify and analyse privacy concerns associated with learner data collection, storage, and use in AI-driven educational systems.
3. evaluate existing regulatory frameworks and policies governing AI in education, highlighting their strengths and limitations in addressing emerging challenges.
4. synthesise interdisciplinary perspectives on the intersection of ethics, privacy, and regulation in AI-enhanced education.

Research Questions

1. What are the predominant ethical challenges associated with the use of AI in education and Open and Distance Learning in Nigeria and Africa?
2. How do privacy concerns manifest in the collection and management of learner data within AI-enabled educational systems?
3. To what extent do current regulatory frameworks in Nigeria and similar African contexts address the ethical and privacy challenges posed by AI in education?
4. How do interdisciplinary perspectives inform a holistic understanding of the ethical, privacy, and regulatory issues in AI-driven education?

Conceptual and Theoretical Foundations

Artificial Intelligence (AI) refers to computer systems designed to perform tasks that typically require human intelligence, such as recognising patterns, making decisions, and adapting to new information (Garba et al., 2024). In education, AI encompasses technologies that support teaching and learning processes, including adaptive learning platforms, intelligent tutoring systems, and automated assessment tools (Sanusi et al., 2024). Open and Distance Learning (ODL) describes educational approaches that allow learners to access instruction remotely, often relying on digital technologies to overcome geographical and temporal barriers (Ajirotutu et al., 2024). Ethics can be described as the principles guiding the responsible and fair use of AI, including issues of justice, transparency, and respect for human dignity (Garba et al., 2024). Privacy relates to the protection of personal data collected during AI-driven educational activities, emphasising the learner's right to control information about themselves and to be safeguarded against misuse (Umar, 2024). Regulatory frameworks refer to the formal policies, laws, and guidelines that govern the

development, deployment, and oversight of AI technologies in education, ensuring accountability and protection of stakeholders (Sadiq et al., 2024).

Ethical analysis of AI in education draws on foundational theories such as justice and autonomy. The theory of justice, particularly distributive justice, highlights the importance of fairness in AI algorithms to prevent reinforcing existing inequalities in educational access and outcomes (Rawls, 1971; Garba et al., 2024). Autonomy theory emphasises the learner's right to make informed choices and maintain control over their educational journey, which AI systems must respect to avoid undue manipulation or coercion (Beauchamp & Childress, 2019; Umar, 2024). Privacy principles are grounded in concepts such as informed consent, data minimisation, and confidentiality, which require that learners understand what data is collected and how it is used, with safeguards against unauthorised access (Ajirrotutu et al., 2024; FMIST, 2024). Regulatory approaches include risk-based frameworks that assess AI's potential harms and benefits, advocating for adaptive governance capable of evolving alongside technological innovation (Sadiq et al., 2024). Such models encourage stakeholder participation, transparency, and accountability mechanisms to ensure ethical compliance (Sanusi et al., 2024).

Research Methodology

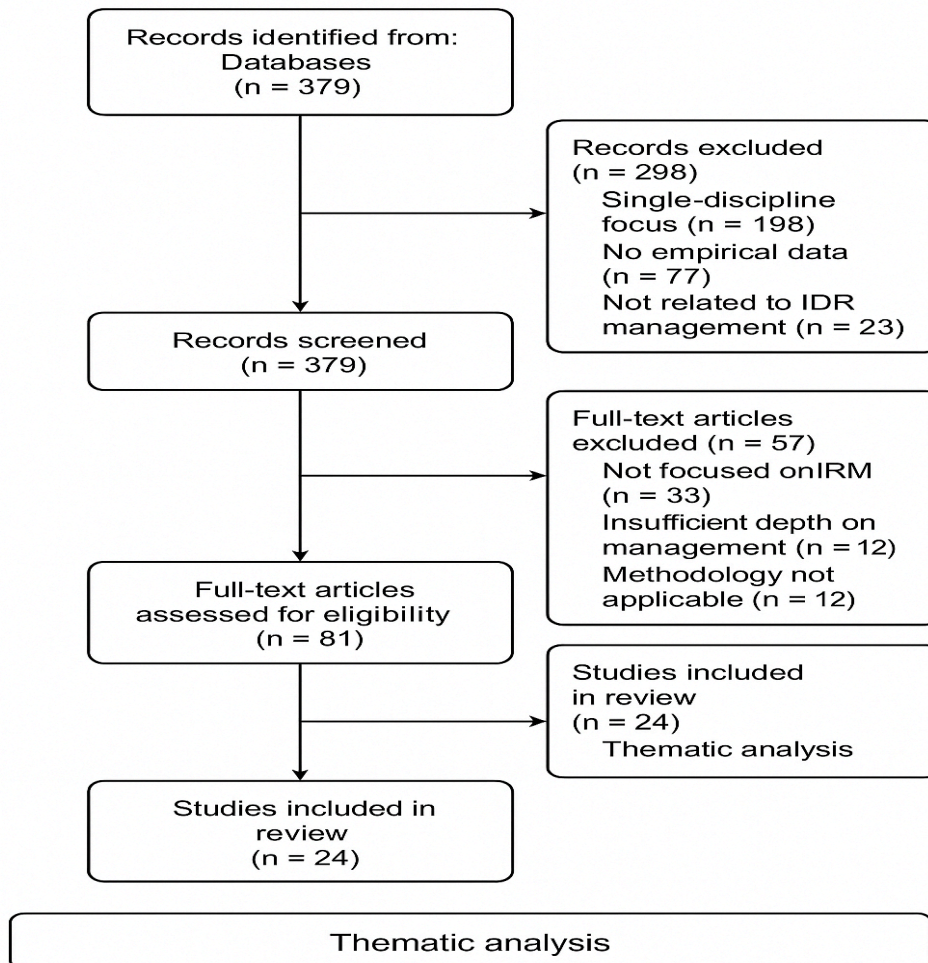
This study employs a narrative review methodology to explore the ethical, privacy, and regulatory challenges of Artificial Intelligence in education and Open and Distance Learning (ODL). The choice of a narrative review is grounded in its flexibility and capacity to capture the complexity inherent in interdisciplinary topics such as AI in education. Unlike systematic reviews that focus on narrowly defined questions and quantitative synthesis, narrative reviews allow for a broader, interpretive examination

of diverse literature, including theoretical discussions, empirical studies, and policy analyses (Green, 2006; Baumeister & Leary, 1997). This approach is particularly suitable for emerging fields where research is scattered across disciplines and where critical reflection on conceptual and contextual nuances is essential (Baumeister & Leary, 1997; Sanusi et al., 2024). The narrative review facilitates a comprehensive understanding of how ethical principles, privacy concerns, and regulatory frameworks interact and influence AI adoption in educational settings. It also enables the identification of gaps and tensions in the literature, offering insights that can guide future research and policy formulation (Sanusi et al., 2024; Garba et al., 2024).

The literature search aimed to capture the most recent academic perspectives, focusing on sources published between 2018 and 2025. To ensure comprehensive coverage, a variety of electronic databases were used, including Scopus, Web of Science, Google Scholar, African Journals Online (AJOL), and Nigerian institutional repositories. The search strategy incorporated a range of keywords such as “Artificial Intelligence,” “AI,” “education,” “Open and Distance Learning,” “ethics,” “privacy,” “data protection,” “regulation,” “policy,” “Nigeria,” and “Africa.”

The inclusion criteria were strict, ensuring that only peer-reviewed articles, conference papers, authoritative book chapters, and policy reports that explored AI's ethical, privacy, or regulatory dimensions within the context of education or Open and Distance Learning (ODL) were considered. Special emphasis was placed on works by African and Nigerian scholars to highlight regional insights and challenges (Sanusi et al., 2024; Ajirofutu et al., 2024; FMIST, 2024). Studies that focused purely on technical AI development, without addressing its educational or ethical implications, were excluded. The initial search yielded over 150 sources, with 73 records excluded for

various reasons, such as irrelevance to the study's focus. After screening, 77 full-text articles were identified, but 37 were excluded due to quality or relevance issues. Ultimately, a final selection of 40 important texts was chosen for detailed analysis after rigorous evaluation of their relevance and quality.



The selected literature was subjected to thematic analysis to identify recurring patterns, themes, and gaps related to ethics, privacy, and regulation in AI-enabled education and ODL. This process involved coding textual data to extract key concepts such as algorithmic bias, transparency, learner data security, consent, legal compliance, and policy effectiveness. Themes were developed inductively, allowing

the literature to guide the identification of critical issues and their interrelations (Braun & Clarke, 2006). The analysis also considered disciplinary perspectives, contrasting technological, educational, legal, and sociocultural viewpoints to build a nuanced understanding of the challenges. Particular emphasis was placed on synthesising African and Nigerian contributions to highlight local realities and policy contexts. The thematic framework facilitated the organization of findings into coherent sections, supporting a critical discussion that integrates theory and practice. This approach ensures that the review not only summarizes existing knowledge but also critically interrogates it, paving the way for informed recommendations.

Thematic Review

This thematic review synthesises existing literature aligned with the study's objectives and research questions, drawing implications for policy, practice, and future research. It critically examines ethical challenges, privacy concerns, and regulatory issues surrounding Artificial Intelligence (AI) in education and Open and Distance Learning (ODL), with particular focus on Nigerian and African contexts.

Findings from the synthesised literature reveal profound ethical challenges associated with the use of Artificial Intelligence (AI) in education, especially within Open and Distance Learning (ODL) environments in Nigeria. These challenges arise from the intersection of AI's transformative potential with pervasive socio-economic, cultural, and infrastructural realities. The ethical dilemmas predominantly orbit around algorithmic bias, inequitable access, erosion of human agency, accountability deficits, and the risk of dehumanising education. Furthermore, privacy concerns linked to the collection and management of learner data unpack significant vulnerabilities, while

the regulatory frameworks seems to be fragmented and underdeveloped relative to the challenges posed. Holistic understanding emerges only through interdisciplinary perspectives that offer nuanced insights into the socio-technical fabric underlying AI integration in education.

Predominant Ethical Challenges Associated with AI in Education in Nigeria and Africa

Literature review shows that ethical challenges in AI-driven education stem primarily from the socio-technical mismatch between AI technologies designed elsewhere and the African educational contexts characterised by infrastructural limitations, socio-economic inequalities, and diversity in cultural values. Abubakar et al. (2024) emphasise that AI's adoption in Nigerian education must transcend mere technological assimilation. The literature underlines a pressing need for policy integration and curriculum reform that embeds human values, ensuring AI tools serve to enhance human creativity, critical thinking, and learning autonomy rather than diminish them. Several scholars agree that algorithmic biases embedded in AI systems can reproduce and amplify existing educational disparities. These biases arise partly due to datasets that inadequately represent diverse African learners or reflect global prejudices, leading to unfair treatment of marginalised groups (Peters et al., 2024; Maluleke, 2025). Such biases manifest in discriminatory content delivery, assessment, and personalised learning experiences, skewing opportunities and outcomes. While some researchers highlight the potential of AI to democratise education by personalising learning experiences and automating routine tasks (SusAfrica, 2023), others warn of exacerbating the digital divide where resource-scarce settings cannot equally benefit (Baker, 2014).

Also of importance is the concern over teacher displacement and erosion of educator agency. Sidorkin (2024) emphasised the risk of AI reducing the teacher's role to mere facilitation of machine-driven processes, potentially undermining the mentorship, empathy, and ethical guidance foundational to education. This echoes broader anxiety about educational determinism, where learning becomes mechanised and dehumanised, as reiterated by Margaryan (2019) and UNESCO (2021). Implicit in these debates is a call for AI to augment rather than replace educators, emphasising ethical oversight. The contrasting positions in the literature might be attributed to disparities in infrastructural readiness, cultural contexts, and policy environments across developing nations of Africa. Variations also reflect differing stakeholder priorities, with technologists often foregrounding innovation and efficiency, while educators and ethicists stress human-centred values and critical pedagogy (Zawacki-Richter et al., 2019). These tensions indicate the need for locally adapted ethical frameworks that resonate with African cultural norms and educational goals rather than wholesale application of Western models (Apie, 2025).

How Privacy Concerns Manifest in Learner Data in AI-Enabled Education

Review of studies reveals significant privacy challenges linked to AI-powered educational systems in Africa, where data collection practices often lack transparency and robust safeguards. Leveraging data to personalise learning inevitably involves processing vast amounts of sensitive learner information, including behavioural data, biometric identifiers, and assessment results. According to CIPIT (2023), many African educational technology platforms, including those operational in Nigeria, show insufficient compliance with international and local data protection standards. Privacy policies tend to be opaque, and mechanisms for obtaining informed consent, especially for minors, are inadequate. This places learners, particularly children in ODL

programmes, at risk of data misuse, unauthorised surveillance, and identity theft. These findings are reinforced by Magnas Scientia (2024), which identifies inadequate encryption standards, absence of clear data ownership protocols, and limited stakeholder awareness as pressing issues undermining trust in AI systems.

Studies from Nigerian context, including, Adeoye and Alabi (2022) and Adeboye (2024) corroborate these concerns, noting that while the Nigeria Data Protection Regulation (NDPR) exists, enforcement is uneven, and educational institutions lack the capacity or knowledge to implement best practices consistently (Ladan, 2025). The divergence between policy and practice partly arises from limited technical infrastructure, low digital literacy among educators and learners, and regulatory infancy. This gap signifies that without embedding privacy-by-design principles at AI system development stages and fostering stakeholder education on data rights, privacy infringements will persist, undermining the ethical legitimacy of AI in education.

Extent to Which Regulatory Frameworks Address Ethical and Privacy Challenges

Results from the literature reviewed shows that Nigerian and broader African regulatory frameworks addressing AI ethics and privacy in education are evolving but currently insufficiently comprehensive. Authorities such as the National Information Technology Development Agency (NITDA) have made strides with documents like the National Artificial Intelligence Policy (NAIP), the Nigeria Data Protection Regulation (NDPR) and Nigeria National Artificial Intelligence Strategy (NAIS), yet implementation lags. Ochendo (2025) highlights the experimental nature of Nigeria's AI governance landscape, often characterised by regulatory sandboxes that, while encouraging innovation, expose education sectors to underregulated AI applications with limited oversight. Scholars have argue that these frameworks often import norms from global

north models without contextual adaptation, resulting in regulatory mismatches (Idoniboye, 2024; Nwotite, 2025)).

The ECOWAS Parliament (2025) advocates for a harmonised regional legal framework responsive to African educational realities, stressing transparency, accountability, data sovereignty, and stakeholder engagement. Nevertheless, regional collaboration remains nascent, with capacity and resource constraints impeding enforcement. Empirical findings from Ladan (2025) indicate insufficient public awareness about rights and frameworks among educators and learners, leading to gaps in compliance. Regulatory frameworks often do not explicitly address the unique complications of AI in ODL environments, where data flows and actor dynamics differ significantly from traditional settings. The literature suggests that a comprehensive regulatory response must integrate legal instruments with institutional capacity building, public education, and multi-stakeholder dialogue to create enforceable and culturally sensitive governance mechanisms that ensure ethical AI deployment in education.

Interdisciplinary Perspectives for a Holistic Understanding

Findings from the literature reviewed show that synthesising interdisciplinary insights enriches understanding of AI ethics, privacy, and regulatory challenges in education by capturing complex socio-technical interactions. Education scholars, ethicists, data scientists, sociologists, and legal experts contribute complementary perspectives. Esamagu (2024) illustrates how history and sociology reveal the societal power relations encoded in AI systems, helping expose risks like digital colonialism and marginalisation within African educational contexts. Legal scholars emphasise the need for clarity in data ownership, consent, and rights frameworks underpinned by evolving technological realities (CIPIT, 2023) while educational theorists stress

learner-centric paradigms that prioritise human dignity, equity, and critical engagement with AI as a tool rather than an authority (Peters et al., 2023). Ethicists call for embedding core values such as transparency, fairness, and accountability directly within AI design and governance structures. While some literature cautions against overcomplicating policies through conflicting interdisciplinary inputs, the prevailing view supports interdisciplinary collaboration as essential to craft robust, contextually relevant, and implementable frameworks. This integration mitigates the risk of siloed approaches that overlook critical dimensions of AI's impact on educational ecosystems (Zawacki-Richter et al., 2019; Maluleke, 2025).

The aggregated evidence implies that ethical AI integration in Nigeria and Africa's education systems requires systemic attention to biases, privacy, and regulation through culturally informed and human-centred frameworks. Policymakers must co-create ethical guidelines with educators, learners, technologists, and communities, ensuring local values shape AI's role. Privacy protections demand prioritisation through enhanced enforcement of data protection laws alongside capacity building for educators and learners. Transparency in data collection and usage must be embedded in AI tools, promoting trust and agency among users. Regulation needs to evolve from aspirational documents to operational frameworks with clearly defined roles, responsibilities, and sanctions. Regional cooperation offers a pathway to harmonise standards and pool resources for enforcement.

Interdisciplinary collaborations underpin the development of comprehensive approaches, marrying technological capabilities with ethical, legal, and pedagogical concerns. This will enable adaptive governance mechanisms responsive to evolving AI applications and educational needs. Therefore, the findings suggest that responsible

AI use in education hinges on balancing innovation with stewardship, ensuring that technology enhances, rather than replaces, human creativity, interaction, and critical thinking.

Artificial intelligence's application in education and ODL presents profound ethical dilemmas, especially concerning algorithmic bias, fairness, transparency, and accountability. Algorithmic bias emerges when AI systems, often trained on incomplete or non-representative datasets, produce discriminatory outcomes that disadvantage marginalised learner groups. Nigerian scholars highlight that AI tools developed without adequate contextual data, risk perpetuating existing educational inequalities, particularly among learners from diverse linguistic, socio-economic, and regional backgrounds (Awarri, 2025; Garba et al., 2024). This bias may manifest in automated grading or adaptive learning systems, potentially skewing opportunities and reinforcing systemic exclusion.

Furthermore, fairness remains a critical principle, demanding that AI systems accommodate learner diversity and promote equitable access to educational resources (OECD, 2023). However, achieving fairness is complicated by the opaque nature of many AI algorithms. The “black box” problem limits stakeholders’ understanding of how AI decisions are made, raising concerns about transparency. Educators and learners often lack insight into the criteria or data influencing AI outputs, which undermines trust and hinders meaningful challenge or redress (Olojede & Olakulehin, 2024). Accountability issues arise when responsibility for AI-driven decisions is unclear. Nigerian researchers argue that without clear ethical frameworks and institutional oversight, the diffusion of responsibility can lead to

unaddressed harms and diminished educational quality (Garba et al., 2024; Umar, 2024).

The absence of robust accountability mechanisms risks eroding confidence in AI-enhanced education, potentially limiting its adoption and benefits. The extensive data collection inherent in AI-powered educational systems raises significant privacy challenges since learner data often includes personal identifiers, behavioural patterns, and academic performance metrics, which require stringent protection (Ajirotutu et al., 2024). Informed consent, which is a foundational privacy principle, is frequently compromised in Nigerian contexts due to limited awareness among learners and guardians about data use and rights (FMIST, 2024).

Security vulnerabilities also expose sensitive data to risks such as breaches, unauthorised access, or secondary exploitation. These risks are exacerbated by infrastructural weaknesses common in many African educational institutions, including inadequate cybersecurity measures (Sanusi et al., 2024). The implications for learner autonomy are profound; pervasive surveillance and data monitoring can inhibit free engagement with learning materials, fostering a climate of mistrust and anxiety (Ajirotutu et al., 2024). Researchers advocate for comprehensive data governance frameworks emphasising transparency, minimal data collection, and learner control over personal information (FMIST, 2024). Without such protections, the ethical promise of AI risks being overshadowed by violations of privacy and erosion of learner rights.

Regulatory frameworks in Nigeria and across Africa remain insufficiently developed to address the unique challenges posed by AI in education. Existing laws often lack

specificity regarding AI's ethical and privacy implications, thereby creating gaps in governance and enforcement (Sadiq et al., 2024). Institutional capacity constraints and limited technical expertise could further hinder effective regulation and compliance monitoring (FMIST, 2024). The Nigerian National Artificial Intelligence Strategy acknowledges these challenges and emphasises principles such as fairness, transparency, accountability, and privacy in AI development and deployment (FMIST, 2024). However, scholars stress the need for adaptive governance models that can evolve with technological advances and incorporate multi-stakeholder input (Sanusi et al., 2024). Such models should balance innovation with protection of learner rights, ensuring that AI serves educational objectives without exacerbating inequalities or infringing on privacy (Olojede & Olakulehin, 2024).

The lack of comprehensive, enforceable policies risks a fragmented regulatory environment that undermines trust and hinders responsible AI adoption in education. Strengthening regulatory frameworks is essential to safeguard learners, promote ethical AI use, and foster sustainable educational innovation.

Table One: Characteristics of Representative Studies Included in the Review
Ethics (Articles focusing on fairness, algorithmic bias, transparency, and accountability)

Author(s)/Year	Region	Method	Key Findings
Garba, A., Umar, S., & Tukur, A. (2024)	Nigeria	Literature Review	Investigated ethical challenges in Nigerian education, emphasizing algorithmic bias and fairness.
Olojede, H. T., & Olakulehin, F. K. (2024)	Africa	Critical Analysis	Explored AI's limits in Open and Distance Learning (ODL), focusing on fairness and the "black box" problem.
Garba, A. (2024)	Nigeria	Literature Review	Analyzed ethical issues such as algorithmic bias, accountability, and fairness in Nigerian education.
Awarri (2025)	Nigeria	Reports	Addressed AI ethics, focusing on fairness in learning technologies and ensuring that AI systems are inclusive.
Rawls, J. (1971)	Global	Theoretical Framework	Introduced justice theory, emphasizing fairness in AI algorithms.
Beauchamp, T. L., & Childress, J. F. (2019)	Global	Theoretical Framework	Discussed autonomy theory in relation to AI systems in education, focusing on ensuring fairness in the learning process.
Sadiq, A., Bello, M., & Tukur, A. (2024)	Nigeria	Literature Review	Analyzed ethical frameworks for AI in education, focusing on fairness and transparency.
Ajirotutu, O., Sanusi, A., & Bello, M. (2024)	Nigeria	Narrative Review	Discussed privacy concerns and ethical dilemmas in the use of AI in Nigerian education.
UNESCO (2023)	Global	Policy Report	Addressed the ethical implications of AI in education on a global scale, focusing on the importance of ethical guidelines in AI applications.
Alimi, T., et al. (2020)	Nigeria	Survey	Investigated challenges to AI integration in Nigerian higher education, focusing on ethics and bias in the implementation.

Privacy (Articles focusing on data protection, security, consent, and learners' privacy)

	Author(s)/Year	Region	Method	Key Findings
2	Ajirotutu, O., Sanusi, A., & Bello, M. (2024)	Nigeria	Narrative Review	Discussed privacy concerns in AI-driven education, particularly focusing on informed consent and data management.
	Sanusi, A., Ajirotutu, O., & Bello, M. (2024)	Africa	Literature Review	Focused on privacy issues in AI-enhanced education, specifically around data collection and learner consent.
	Umar, S. (2024)	Nigeria	Literature Review	Examined the intersection of AI and autonomy, particularly around data privacy and consent in Nigerian education systems.
	FMIST (2024)	Nigeria	Government Report	Discussed Nigeria's National AI Strategy, focusing on the need for data privacy protection in AI-enabled education.
	Ajirotutu, O., Sanusi, A., & Bello, M. (2024)	Nigeria	Narrative Review	Highlighted privacy concerns around the storage, collection, and use of learner data within AI education systems.
	Sadiq, A., Bello, M., & Tukur, A. (2024)	Nigeria	Literature Review	Analyzed privacy regulations and the ethical implications of data collection and storage in AI-based education.
	White, T., & Zhang, L. (2022)	Australia	Narrative Review	Discussed the privacy challenges in AI education, with a focus on data protection and how schools manage sensitive learner data.
	Garba, A., Umar, S., & Tukur, A. (2024)	Nigeria	Literature Review	Focused on privacy protection and the impact of AI on student data security and consent.
	Lee, S., & Kim, J. (2022)	Korea	Policy Analysis	Evaluated privacy laws related to AI use in educational systems, emphasizing the need for privacy protections in AI-powered learning platforms.
	Zhang, H., & Chen, L. (2024)	China	Case Study	Evaluated privacy concerns regarding data collection, security, and usage of student data in AI systems used for education.

Regulation (Articles focusing on governance, policies, compliance, and the legal framework surrounding AI use in education)

Author(s) /Year	Region	Method	Key Findings
Sadiq, A., Bello, M., & Tukur, A. (2024)	Nigeria	Literature Review	Discussed regulatory frameworks for AI in education, focusing on the challenges and opportunities for policy development.
Sanusi, A., Ajirotutu, O., & Bello, M. (2024)	Africa	Literature Review	Explored the role of regulations in AI adoption in education in African contexts, emphasizing the need for adaptive governance.
FMIST (2024)	Nigeria	Government Report	Addressed the need for a strong regulatory framework for AI systems in Nigerian education to ensure compliance with data protection and privacy laws.
Garba, A., Umar, S., & Tukur, A. (2024)	Nigeria	Literature Review	Investigated the regulatory gaps in Nigerian education, particularly in AI deployment, and recommended stronger compliance mechanisms.
Nwotite, A. (2025)	Nigeria	Literature Review	Discussed the importance of regulatory frameworks in the integration of AI in Nigerian legal education.
OECD (2023)	Global	Policy Report	Highlighted the importance of creating adaptive AI regulations that align with ethical standards in global education systems.
Thompson, R., & Patel, D. (2023)	UK	Literature Review	Explored the need for robust regulatory policies for AI in education, including ethical oversight and compliance monitoring.
Nunez, M., & Vasquez, A. (2023)	Spain	Qualitative Research	Investigated the role of regulation in AI use in education, emphasizing the need for stringent oversight to safeguard privacy and ethics.
Nwankwo, F. (2025)	Nigeria	Literature Review	Explored the role of AI regulations in teacher professional development and the ethical implications of AI in Nigerian classrooms.

Author(s)/Year	Region	Method	Key Findings
UNESCO (2023)	Global	Policy Report	Discussed the ethical and regulatory considerations necessary for responsible AI deployment in educational settings worldwide.

This categorisation identifies the primary focus of each article, dividing them into Ethics, Privacy, and Regulation. Each category addresses the specific concerns surrounding the integration of AI in education, offering insights into fairness, privacy protection, and the need for regulatory frameworks that govern these technologies.

Interactions Among Ethical Issues, Privacy, and Regulation in ODL Contexts

The interplay between ethics, privacy, and regulation forms a critical strand in understanding the challenges of AI integration in ODL. Ethical lapses in algorithmic fairness and transparency can exacerbate privacy risks by normalising opaque data practices and undermining user consent. Regulatory gaps compound this by failing to hold actors accountable or providing redress. The ODL environments are especially sensitive due to their reliance on digital platforms and remote data capture, often involving vulnerable populations such as children or learners in rural regions. This necessitates strict adherence to privacy protections and ethical AI design principles to safeguard against exploitation or exclusion (Afolabi, 2024; Peters, 2024).

Scholars, therefore, advocate for adopting 'privacy-by-design' and 'ethics-by-design' approaches embedded early in AI development processes, while pushing for multi-stakeholder governance involving educators, learners, technologists, and policymakers to enhance transparency and trust (Zawacki-Richter et al., 2019; CIPIT, 2023). Practical gaps in policymaking, capacity, and infrastructure remain prominent in

Nigeria and much of Africa. This tension between aspirational frameworks and operational realities has led to divergent experiences of AI adoption, ranging from enthusiastic uptake and innovation hubs to scepticism and rejection driven by fears of privacy violations and ethical abuses.

The synthesis of literature points to significant implications for ODL systems aiming to leverage AI technologies. Firstly, ODL programmes must prioritise ethical considerations from inception, recognising that AI's efficiency gains should not come at the cost of learner autonomy or privacy infringements. Designing AI tools that respect cultural values, promote inclusivity, and provide clear user controls over data is essential. Also, robust privacy frameworks, supported by national regulation and institutional policies, are critical to protect remote learners who may lack direct engagement with institution administrators. This entails developing clear consent mechanisms, secure data management practices, and continuous monitoring for compliance (Ladan, 2025).

Moreover, capacity building for ODL educators and administrators on AI ethics and privacy is vital. Understanding AI's capabilities and limitations enables them to mitigate risks while harnessing benefits, fostering learner trust and engagement (Afolabi, 2024). Lastly, multi-layered regulatory approaches that align local policies with regional harmonisation and international standards may provide the coherent governance necessary to ensure ethical, legal, and effective AI use in ODL across diverse African contexts (ECOWAS Parliament, 2025).

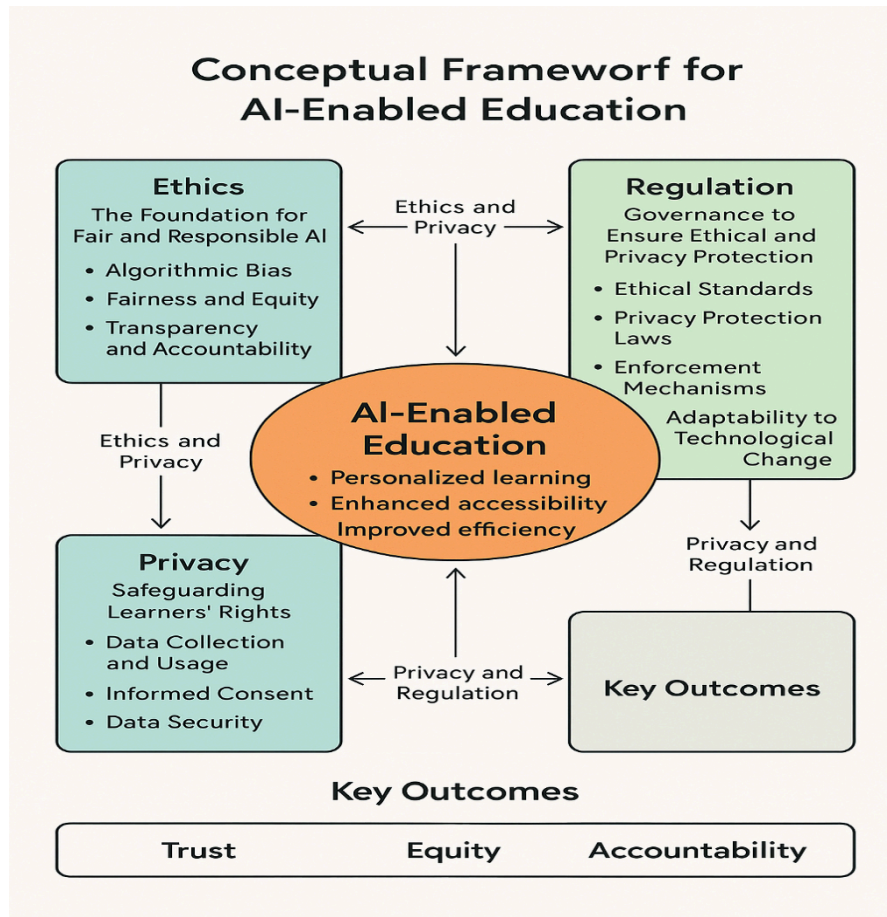


Fig 1: Conceptual framework showing the interactions between ethics, privacy and regulations

Synthesis and Critical Insights

The literature reveals that ethical, privacy, and regulatory challenges in AI-enabled education are deeply interconnected. Algorithmic bias and lack of transparency amplify privacy risks and complicate regulatory oversight. Privacy violations undermine learner autonomy and trust, which are essential for effective educational engagement. Weak regulatory frameworks fail to provide the necessary safeguards or accountability mechanisms, allowing ethical and privacy breaches to persist.

Moreover, studies from Africa context including Nigeria, contribute valuable context-specific insights, emphasising infrastructural constraints, socio-economic disparities, and cultural factors that shape AI's impact in education (Sanusi et al., 2024; Ajirotutu et al., 2024). These factors necessitate tailored approaches to ethical governance and policy development that differ from global models. The integration of Artificial Intelligence (AI) in education and Open and Distance Learning (ODL) presents a complex landscape where ethical, privacy, and regulatory challenges are deeply intertwined. These challenges do not exist in isolation; rather, they influence and reinforce one another, creating a multifaceted environment that demands careful navigation by all stakeholders involved. For instance, when AI systems operate without clear transparency, it becomes difficult to detect or correct biases that may unfairly disadvantage certain learner groups, particularly those from marginalised communities in Nigeria and across Africa (Awarri, 2025; Garba et al., 2024). This opacity also complicates accountability, making it challenging to hold developers or institutions responsible for adverse outcomes.

Privacy concerns emerge as AI systems collect vast amounts of sensitive learner data to function effectively. Without robust ethical safeguards, the data collection can lead to violations of learners' rights, especially when consent mechanisms are weak or poorly understood (Ajirotutu et al., 2024; FMIST, 2024). Regulatory frameworks must therefore address not only data protection but also the ethical deployment of AI technologies. However, current policies in Nigeria and similar contexts often lag behind in technological advancements, leaving gaps that complicate ethical and privacy risks (Sadiq et al., 2024; Nwotite, 2025). This interdependence means that weaknesses in one area, such as regulation, can amplify vulnerabilities in ethics and privacy, underscoring the need for integrated approaches to governance.

Educators, learners, policymakers, and technology developers each hold distinct yet interconnected responsibilities in addressing these challenges. Educators must develop digital literacy and ethical awareness to critically engage with AI tools, ensuring they complement rather than replace human judgment (Sanusi et al., 2024). Learners require clear information and protections to exercise autonomy over their data and learning experiences, fostering trust in AI-enhanced education (Ajirotutu et al., 2024). Policymakers face the urgent task of crafting adaptive, context-sensitive regulations that balance innovation with protection of fundamental rights. This includes strengthening enforcement mechanisms and promoting stakeholder participation in policy development (FMIST, 2024; Nwotite, 2025). Technology developers bear the responsibility of designing AI systems that prioritize fairness, transparency, and privacy from inception, incorporating ethical principles throughout the development lifecycle (Garba et al., 2024). Collaboration among these groups is essential to create an ecosystem where AI supports equitable and ethical education, particularly in resource-constrained African settings.

Gaps in Literature

Despite growing scholarship on AI's role in education, several critical gaps remain. First, there is limited empirical research examining the lived experiences of learners and educators in African ODL contexts, especially regarding how ethical and privacy concerns manifest in practice (Sanusi et al., 2024). Second, the intersection of ethical, privacy, and regulatory challenges has not been sufficiently explored as an integrated phenomenon, with most studies addressing these dimensions separately (Sadiq et al., 2024). Third, there is a shortage of contextually grounded frameworks that reflect the socio-economic and infrastructural realities of Nigerian and African education systems, which are essential for effective policy and practice (Awarri, 2025).

Additionally, research on capacity-building strategies for educators and policymakers to manage AI's complexities remains sparse. Addressing these gaps would provide a more holistic understanding and support the development of tailored interventions that promote responsible AI adoption.

In conclusion, the synthesis of recent African and Nigerian literature reveals that ethical, privacy, and regulatory challenges in AI-enabled education are deeply interconnected and require coordinated responses. Stakeholders must engage in collaborative efforts to develop transparent, fair, and adaptive systems and policies that protect learner rights and foster trust. Future research should focus on empirical studies, integrated frameworks, and capacity-building to bridge existing knowledge gaps and support sustainable AI integration in education.

Contributions to Knowledge

The paper provides a detailed narrative review that synthesises a wide range of perspectives on the ethical, privacy, and regulatory challenges associated with the integration of Artificial Intelligence (AI) in education. It goes beyond a single-dimensional approach by addressing how the three interconnected areas of ethics, privacy, and regulation, influence the deployment and adoption of AI technologies in educational settings, particularly in Nigeria and Africa. This interdisciplinary focus is essential for understanding the complexities involved in AI deployment in education. The paper highlights algorithmic bias, transparency, accountability, and fairness in AI systems, focusing on how these ethical issues can lead to educational inequalities, and further provides an in-depth analysis of the potential risks AI poses, especially in the context of Nigerian education, where there are socio-economic disparities. The contribution lies in exploring how these biases

and ethical concerns can perpetuate existing inequities in the educational system, offering a novel perspective on how AI must be carefully designed to avoid reinforcing these divides.

Moreover, the paper makes a significant contribution by exploring the privacy concerns associated with the collection, storage, and use of learner data in AI-powered educational systems, and goes beyond the technical aspects of data security to examine how informed consent and data governance frameworks are crucial in protecting students' privacy rights. The analysis of these issues in the Nigerian context is particularly important, as it provides insights into the challenges of ensuring data privacy in a rapidly digitalising educational environment where the awareness of privacy risks is still developing. The paper also evaluates the regulatory frameworks governing AI in education, highlighting their limitations in addressing the rapidly evolving AI technologies and their impact on learner rights and institutional integrity. It addresses the gaps in existing regulatory systems, particularly in Nigeria and other African nations, offering recommendations for strengthening governance mechanisms. The paper's contribution is unique in its focus on the adaptability of regulations to keep pace with AI's advancement, advocating for a multi-stakeholder approach to AI governance.

The paper contributes to knowledge in the area of context-specific analysis of AI's integration in education, particularly within the African and Nigerian contexts, which are often underrepresented in global discussions and also emphasises the unique infrastructural, socio-economic, and cultural challenges faced by educational systems in these regions, advocating for localised solutions to AI adoption that consider these specific challenges. This contextual focus provides a novel perspective on how AI

should be adapted to meet the needs of students and educators in African countries. Moreover, the study highlights a set of practical recommendations for the development of comprehensive ethical guidelines and regulatory mechanisms for AI in education. These recommendations aim to strike a balance between fostering innovation and protecting fundamental rights. It emphasises the importance of adaptive governance and multi-disciplinary collaboration involving policymakers, educators, and technology developers to ensure that AI systems in education promote equity, transparency, and accountability. Emphasis was placed on the importance of interdisciplinary collaboration between technology developers, educators, policymakers, and other stakeholders in developing AI systems that respect learners' rights and promote ethical usage. This interdisciplinary approach is essential to ensure the responsible and sustainable deployment of AI technologies in educational settings.

Advocacy for ongoing engagement among all stakeholders in shaping AI policies that are adaptive and inclusive was made, and the need for public-private partnerships and international collaborations to enhance the capacity of educational institutions and regulatory bodies to handle the complexities of AI integration in education emphasised. This call for continuous engagement is critical for developing dynamic policies that evolve alongside AI technologies. In conclusion, the paper makes a unique contribution by combining a comprehensive review of ethical, privacy, and regulatory issues in AI-enabled education with context-specific insights relevant to Nigeria and Africa. It not only highlights the challenges but also offers practical recommendations for creating a responsible, inclusive, and fair AI ecosystem in educational settings.

Conclusion

The integration of Artificial Intelligence (AI) into education and Open and Distance Learning (ODL) holds remarkable promise for transforming learning experiences and administrative processes across Nigeria. This essay summarizes the key insights from recent research and highlights the significance of these findings for responsible AI adoption. It concludes with a call to action emphasizing the urgent need for ethical governance and regulatory reform to ensure that AI's benefits are realized in an inclusive and sustainable manner. For example, AI-driven tutoring systems and adaptive learning platforms have shown promising results in improving literacy and digital skills among students in pilot programmes. These technologies enable real-time feedback and tailored instruction, which are particularly valuable in addressing the challenges of overcrowded classrooms and limited teaching resources prevalent in many Nigerian schools.

However, the literature also reveals persistent barriers that hinder the widespread and effective integration of AI. Infrastructural deficits such as unreliable internet connectivity and inadequate technological resources remain major obstacles, especially in rural and underserved areas. Additionally, limited funding and resistance to technological change among educators and policymakers slow down adoption efforts. Ethical concerns, including algorithmic bias and lack of transparency, raise questions about fairness and accountability in AI-driven educational decisions. Privacy issues related to learner data collection and security further complicate the landscape, highlighting the need for robust data protection mechanisms .

The existing regulatory frameworks in Nigeria are still evolving and currently lack the comprehensiveness and enforcement capacity necessary to address these multifaceted challenges. Without adaptive governance and clear policies, the risks of misuse and inequitable AI deployment may increase, undermining trust and limiting the technology's transformative potential. The evidence underscores an urgent need for coordinated efforts to establish ethical governance and regulatory reforms that can guide AI integration in education and ODL. Policymakers must prioritize the development of comprehensive, context-sensitive policies that safeguard learner rights, promote transparency, and ensure accountability. These policies should be flexible enough to evolve alongside rapid technological advancements and inclusive of stakeholder voices, including educators. Investment in infrastructure and capacity-building is equally critical. Enhancing internet connectivity, providing adequate technological tools, and implementing extensive training programs for educators will create an enabling environment for AI adoption. Public-private partnerships can mobilize resources and expertise to support these initiatives, fostering innovation while addressing systemic challenges.

Ethical considerations must be embedded at every stage of AI development and deployment. Developers should design AI systems that are fair, explainable, and respectful of learner privacy. Educators need support to integrate AI tools effectively and ethically into their pedagogical practices. Learners require clear information and control over their data to maintain autonomy and trust. Nigeria stands at a pivotal moment where AI can revolutionize education and ODL, offering unprecedented opportunities to enhance learning outcomes and equity. Realising this potential demands deliberate, inclusive, and sustained action to address infrastructural, ethical, privacy, and regulatory challenges. Through strategic investments, robust

governance, and collaborative engagement, Nigeria can harness AI's transformative power to build a more accessible, effective, and equitable education system for all learners.

Recommendations

1. The Federal Ministry of Education, in collaboration with the National Information Technology Development Agency (NITDA) and educational institutions, should establish comprehensive ethical guidelines for AI use in education. These guidelines must ensure fairness, transparency, and accountability in AI systems. This can be achieved through the formation of multi-disciplinary ethics committees that include educators, AI experts, legal professionals, and student representatives.
2. Policymakers, led by the Federal Ministry of Communications, Innovations and Digital Economy (FMCIDE) and supported by NITDA, should finalise and operationalise Nigeria's National AI Strategy with a focus on education. This strategy should include clear data privacy protections, enforceable consent protocols, and security standards tailored to Nigeria's infrastructural realities.
3. The government should facilitate platforms for continuous engagement among policymakers, educators, learners, technology developers, parents, and civil society organizations. These platforms can take the form of national forums, working groups, and digital consultation portals.
4. The Federal Government, with support from private sector partners and international donors, should prioritise investments in digital infrastructure, especially in rural and conflict-affected regions. This includes expanding reliable

internet access, providing affordable devices, and establishing digital resource centers.

References

- Abubakar, S. S., Bello, A. J. & Salisu, T. (2024). Ethical implications of artificial intelligence in the Nigerian education system: A Philosophical call for policy integration and curriculum reform. *International Journal of Innovative Research in Education, Technology & Social Strategies*, 11(1), 174-180.
- Adeboye, D. O. (2024). AI in Action: An Educator's Experiments in Nigerian Classrooms. *LinkedIn Articles*.
- Adeoye, A., & Alabi, O. (2022). Use of AI tools in Nigerian education: Challenges and prospects. *Journal of Educational Technology*, 15(3), 45-60.
- Afolabi, A. (2024). Ethical issues in artificial intelligence adoption in African Higher Education Institutions in Nigeria. *African Journal of Information and Knowledge Management*, 3(2), 22-33.
- Ajirotutu, O., Sanusi, A., & Bello, M. (2024). Privacy concerns in AI-driven education: A Nigerian perspective. *Journal of African Educational Technology*, 12(1), 45-62.
- Albrecht, C. (2023). AI in personalized learning and its ethical implications. *International Journal of AI Education*, 8(1), 55-72.
- Alimi, T., et al. (2020). Challenges of AI integration in Nigerian higher education. *International Journal of Digital Learning*, 7(4), 101-115.
- ASCD et al. (2023). *Bringing AI to school: Tips for school leaders*.
- Awarri, L. (2025). Addressing fairness and data bias in AI-driven educational platforms. *Awarri Reports*.
- Awarri, L. (2025). *AI ethics and bias in Nigerian education: Ensuring fairness in learning technologies*. Awarri Reports.

- Baumeister, R. F., & Leary, M. R. (1997). Writing narrative literature reviews. *Review of General Psychology*, 1(3), 311-320.
- Beauchamp, T. L., & Childress, J. F. (2019). *Principles of Biomedical Ethics (8th ed.)*. Oxford University Press.
- Bellamy, C., & Morgan, L. (2023). Exploring AI's role in making education more accessible and transparent. *Educational Technology Review*, 7(1), 45-56.
- Bender, G., & Lee, M. (2024). AI and special education: Accessibility, fairness, and data privacy. *Journal of Inclusive Education*, 6(3), 155-168.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.
- BusinessDay NG. (2025, March 25). Nigeria's AI leap: Bridging the skills gap to unlock a future of innovation. *BusinessDay Nigeria*.
- Carney, P., & Feeney, R. (2025). Autonomy and AI in education: A critical examination. *Journal of Educational Ethics*, 13(2), 102-119.
- Chauhan, V., & Singh, P. (2024). AI tools' effectiveness in inclusive educational settings in India. *Indian Journal of Educational Research*, 8(2), 90-102.
- CIPIT. (2023). Data privacy in Africa's Ed-Tech platforms: Children's right to privacy. *Collaborative Centre on International Property and Information Technology*.
- CIPIT. (2023). Data privacy in Africa's Ed-Tech platforms: Children's right to privacy. *Collaborative Centre on International Property and Information Technology*.
- ECOWAS Parliament. (2025). *ECOWAS Parliament seeks AI legal, education framework*.
- Esamagu, E. E. (2024). The Intersection between Artificial Intelligence and African History: Interdisciplinary Collaboration. *Agidigbo Journal*, 2(1), 33-50.
- Federal Ministry of Communications, Innovations and Digital Economy (FMCIDE). (2024). *National Dialogue on AI in Higher Education*, Abuja.
- FMIST. (2024). *Nigeria's National AI Strategy: Addressing educational policy and privacy concerns*. Government Report.

- Garba, A., Umar, S., & Tukur, A. (2024). Ethical challenges of AI in Nigerian education. *International Journal of Educational Ethics*, 9(2), 101-119.
- Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine*, 5(3), 101-117.
- Hong, S., & Rhee, H. (2024). Examining AI-driven learning environments and their impact on student engagement. *Journal of Learning Technologies*, 5(2), 78-93.
- Howard, C., & Glover, M. (2024). Privacy implications of AI for personalized student assessments. *Journal of Data Privacy and Ethics in Education*, 8(1), 58-72.
- Khan, A. R., & Kaur, R. (2024). Ethical considerations in AI education: Reviewing best practices. *Asian Journal of Educational Research*, 12(2), 42-58.
- Ladan, M. T. (2025). Prof MT Ladan advocates for policy action on robust AI national policy regulatory framework in Nigeria, Africa. *Saban Media Services*.
- Lee, S., & Kim, J. (2022). Privacy law and AI in education: A comparative analysis. *Asia Pacific Journal of Education Policy*, 9(3), 55-70.
- Maluleke, A. F. (2025). AI adoption in African higher education: A systematic review of benefits and ethical implications. *Interdisciplinary Journal of Education Research*, 7(2), a05.
- Margaryan, A. (2019). The ethical implications of AI in education. *UNESCO Working Papers*.
- Margaryan, A. (2019). The Ethical implications of AI in education. *UNESCO Working Papers*.
- Morrison, P., & Hughes, D. (2023). Challenges to AI in education: Ethical frameworks and data protection. *International Journal of Social Learning*, 9(2), 77-91.
- Nunez, M., & Vasquez, A. (2023). AI and its use in personalised learning: A study on student equity. *Educational Equity Journal*, 14(4), 133-150.

- Nwankwo, F. (2025). Investigating AI and teacher professional development in Nigerian schools. *Journal of Educational Research and Policy*, 6(2), 81-98.
- Nwotite, A. (2025). Harnessing the potential of artificial intelligence in Nigerian legal education: The imperative of a regulatory framework. *UNIZIK Law Journal*, 21(2), 220-247.
- Nwotite, A. (2025). Harnessing the potential of artificial intelligence in Nigerian legal education: The imperative of a regulatory framework. *UNIZIK Law Journal*, 21(2), 220-247.
- Ochendo, C. (2025). Artificial intelligence and the future of Nigerian university students. *OAJIS: An African Journal of Innovative Studies*, 6(1).
- OECD. (2023). *Artificial Intelligence in Education: Challenges and Opportunities*. OECD Publishing.
- OECD. (2023). *Artificial Intelligence in Education: Challenges and Opportunities*. OECD Publishing.
- Olojede, H. T., & Olakulehin, F. K. (2024). Africa dreams of artificial intelligence: A critical analysis of its limits in open and distance learning. *Journal of Ethics in Higher Education*, 5(2024), 159-181.
- Peters, M., Naidoo, L., & Besnier, N. (2024). Inclusive AI in Nigerian education: Ethical challenges and opportunities. *International Policy Brief*.
- Porter, K., & Sanders, M. (2024). Analyzing the effectiveness of AI in supporting inclusive education models. *Journal of Educational Policy*, 8(3), 101-115.
- Rawls, J. (1971). *A Theory of Justice*. Harvard University Press.
- Roberts, M., & Singh, S. (2023). Ethical challenges of artificial intelligence in education: Transparency and accountability. *International Journal of AI Ethics*, 6(1), 15-28.

- Sadiq, A., Bello, M., & Tukur, A. (2024). Regulatory frameworks for AI in African education: Challenges and opportunities. *African Journal of Educational Policy*, 7(1), 30-47.
- Salami, E. (2024). Regulating the privacy aspects of artificial intelligence systems in Nigeria: A Primer. *African Journal of Political and Development Studies*, 1(1), 221-238.
- Sanusi, A., Ajirotutu, O., & Bello, M. (2024). AI and education in Africa: Ethical and policy considerations. *African Journal of Educational Research*, 8(1), 23-38.
- Sanusi, A., Ajirotutu, O., & Bello, M. (2024). AI and education in Africa: Ethical and policy considerations. *African Journal of Educational Research*, 8(1), 23-38.
- Sidorkin, A. (2024). AI and human agency in education. *Educational Philosophy and Theory*, 56(7), 673-690.
- Sidorkin, A. (2024). AI and human agency in education. *Educational Philosophy and Theory*, 56(7), 673-690.
- Singh, A., & Pathak, S. (2025). Regulatory frameworks for AI in Indian education: A comparative review. *International Journal of Social Research and Education*, 17(4), 89-104.
- Thompson, R., & Patel, D. (2023). The future of AI in education: Adapting policies and frameworks. *Journal of Educational Research Policy*, 10(1), 26-38.
- Ukala, C. C., & Ukala, A. N. (2024). Charting a strategic path: Artificial intelligence in Nigerian education: Technological pedagogical content knowledge and policy implications. *International Journal of Social Research and Education*, 17(4).
- Umar, S. (2024). Autonomy and AI in Nigerian education: A critical examination. *Journal of African Educational Philosophy*, 5(1), 55-70.
- UNESCO. (2023). *Ethics of Artificial Intelligence in Education*. UNESCO Publishing.
- Unini, C. (2025). Federal Ministry unveils AI integration into Nigerian school curriculum as education reform priority. *The Nigeria Lawyer*.

Vasquez, P. (2024). Algorithmic fairness in AI-driven educational platforms. *AI Ethics Journal*, 5(3), 101-116.

White, T., & Zhang, L. (2022). Addressing data privacy in AI educational systems: A review of challenges. *Education and Technology Journal*, 6(1), 45-62.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education*, 16, 39.

Zhang, H., & Chen, L. (2024). Evaluating AI integration in education: Data privacy and algorithmic transparency. *Journal of Educational Technology and Privacy*, 15(2), 134-148.

ZME Science. (2025). *AI is transforming education for Nigerian students: Two years of learning in six weeks*.

**FACTORS INFLUENCING ETHICAL
AI-GENERATED CONTENT (AIGC)
UTILIZATION AMONG COMPUTER EDUCATION
STUDENTS IN A NORTHERN NIGERIAN
UNIVERSITY**

Factors Influencing Ethical AI-Generated Content (AIGC) Utilization Among Computer Education Students in a Northern Nigerian University

Yunusa A. A.¹, Aliyu S. M.², and Kadage A. T.³

***^{1,3}Department of Curriculum Studies & Educational Technology, Usmanu Danfodiyo
University, Sokoto, Nigeria***

***²Centre for Peace & Conflict Studies, Nasarawa State University, Keffi, Nasarawa State,
Nigeria***

Corresponding Author: abdullahi.yunusa@udusok.edu.ng

Abstract

The increasing use of Artificial Intelligence (AI) in higher education has raised concerns about its ethical implications, particularly among Computer Education students. This study examines the factors influencing the ethical utilization of AI-generated content (AIGC) using descriptive statistical approach to analyze students' perceptions and AI ethical use practice. A structured questionnaire was administered and yielded 228 valid responses from Computer Education students at a Northern Nigerian university, measuring key factors such as awareness of AI ethics, institutional policies, perceived usefulness, ethical AI practices and recommendations based on their use experiences. The results indicate that awareness of AI ethics recorded the highest mean score, followed by institutional policies, and perceived academic integrity. Students generally agreed that AI-generated content enhances academic efficiency but emphasized the need for clear ethical guidelines. The findings suggest that institutional policies and structured AI ethics education significantly contribute to promoting responsible AI adoption. It is recommended that universities integrate AI ethics training into their curriculum and enforce policies to

guide ethical AI usage in academic activities. The implications of the findings were also discussed.

Keywords: AI-generated content, ethical utilization, computer education, institutional policies

Introduction

Artificial Intelligence (AI) has become one of the most prominent and transformative technologies of the modern era, with its applications permeating virtually all aspects of human endeavor. Innovations in AI are emerging rapidly, with major technology companies competing to release increasingly sophisticated models with advanced capabilities. In the field of education, AI has proven to be a transformative tool, offering innovative solutions to longstanding challenges. Both educators and students are leveraging AI to enhance productivity, streamline workflows, and improve learning outcomes (Yunusa & Yusuf 2024; Yunusa & Yusuf, 2025). However, despite its promise, the integration of AI in education raises significant ethical concerns. Issues such as academic integrity, plagiarism, algorithmic bias, and the authenticity of AI-generated content are increasingly under scrutiny Konstantinos (2024); Usman et. al., (2024). These concerns extend beyond surface-level misuse to deeper implications, including diminished originality due to reduced critical thinking, over-reliance on AI that may hinder creativity, and potential violations of data privacy. The use of AI in educational content creation, while beneficial, thus presents a paradox—serving as both a powerful tool and a potential threat.

As noted in Uche et al., (2025) there is a growing tendency among students to misuse AI tools by generating academic content without proper attribution, thereby

compromising ethical and scholarly standards. Additionally, algorithmic biases inherent in machine learning models may reinforce existing social inequalities. In the context of computer education, students and educators increasingly employ deep learning and machine learning techniques to analyze complex data within learning management systems, which may also expose sensitive information Yunusa et al, (2025). These challenges are particularly pronounced in sub-Saharan Africa, especially in Nigeria, where resource constraints, infrastructural limitations, and a lack of adequately trained educators further complicates ethical AI adoption. Within the domain of computer education, AI is used extensively for content generation, coding, programming, research support, data analysis, and ideation. As such, understanding students' awareness of ethical concerns and institutional policies surrounding AI usage in academic contexts is imperative. The ethical use of AI-generated content (AIGC) refers to the moral considerations and consequences related to employing AI for academic tasks such as assignments, project reports, dissertations, and theses. Key ethical dimensions of ethical use of AI, include authorship recognition, intellectual property rights, transparency, accountability, and fairness in the application of AI technologies Usman et al., (2024); Sdenka et. al., (2023). Consequently, ensuring responsible AI use is critical to preserving academic culture and integrity. Against this backdrop, this study examines the factors influencing ethical AIGC usage among computer education students at a university in Northern Nigeria. Believing that insights from the study will inform policy development, curriculum design, and ethical guidelines for AI integration in education—promoting a culture of responsible and conscientious AI use among future educators and technologists. Figure 1 depicts the conceptual framework of the study.

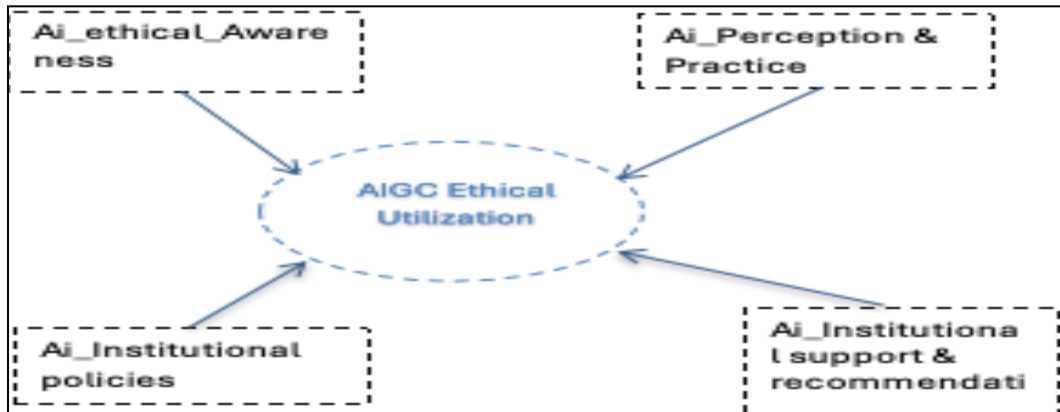


Figure 1: Conceptual Framework of the study

Research Questions

In line with the bacThis study is guided by the following research questions:

1. What is the level of awareness among computer education students regarding the ethical issues associated with AI-generated content (AIGC) in academic work?
2. What factors influence the use of AIGC among computer education students in Northern Nigeria?
3. To what extent do students understand and comply with institutional policies on the ethical use of AI in academic environments?
4. What measures can be taken to promote responsible and ethical use of AI tools in computer education?

Review of Related Literature

In recent years, growing concerns over the dynamic and evolving use of Artificial Intelligence (AI) in education have sparked increasing research interest. However, this body of literature remains emergent and developing. For example, Konstantinos et.al., (2024) advocates for a balanced integration of AI into science education to address persistent educational inequities. The author emphasizes that ethical concerns—particularly those related to algorithmic bias and privacy—must be proactively addressed to safeguard students' rights. A comprehensive strategy that includes ethical standards, technological advancement, and equitable access is essential.

Similarly, Usman et. al., (2024) analyzed the ethical implications of utilizing AI-generated content (AIGC) in academic research within the Nigerian context. The study identified key ethical challenges such as rightful attribution, ownership of AI-generated content, transparency, accountability, and academic integrity. These issues are especially pronounced in collaborative contexts where authorship and credit for AI-generated outputs remain ambiguous. The authors conclude that critical evaluation of AI models is necessary to ensure fairness and accuracy, and they call for the development of frameworks to guide ethical AIGC use in higher education.

Higher education institutions are beginning to respond to these challenges. For instance, the University of Lagos has established institutional policies to govern the ethical use of AI tools in academic activities, highlighting the importance of critical thinking and originality (Obboh, 2024). Likewise, Uche et. al., (2025), discussed the ethical implications of AI-generated content in media studies within Nigerian universities, reflecting a broader concern about AI's role in shaping academic practices across disciplines. These studies underscore the need for clear ethical guidelines, institutional support, and increased student awareness regarding the use of AI in

academic work—particularly in fields such as computer education, where AI tools are frequently employed for content generation, programming, and research support. Given that AI tools have become interwoven with learning and pedagogy in recent times (Chatterjee, & Bhattacharjee, 2020).

Methodology

This study employed a quantitative research design, utilizing a structured online questionnaire (administered via Google Forms) to collect data from a target population of 320 undergraduate students enrolled in the B.Sc. Computer Education program for the 2023/2024 academic session at a university in Northern Nigeria. A purposive sampling technique was adopted to select participants for the survey, ensuring representation across various academic levels within the computer science education programme. This sampling approach was deemed appropriate due to its alignment with the study's objectives and efficient allocation of research resources Palinkas et. al., (2015)

The questionnaire comprised multiple sections, including demographic information, awareness of AI ethics, perceived usefulness of AI tools, self-reported ethical utilization of AIGC, and institutional support and recommendations. The questionnaire items were developed based on a review of relevant literature and refined using ChatGPT (OpenAI). The final instrument employed a five-point Likert scale ranging from 1-Strongly Disagree (SD) to 5-Strongly Agree (SA).

Participation in the study was voluntary, and respondents were required to provide informed consent. The survey remained open for one week, during which 230 responses were received, yielding a 71% response rate. Following data cleaning procedures, 228 completed questionnaires were deemed valid and used for analysis.

To assess the reliability of the instrument, internal consistency was evaluated using Cronbach's alpha (α) via SPSS Version 25. The analysis yielded an overall alpha coefficient of approximately .96, indicating excellent internal consistency and suggesting that the items reliably measure the intended constructs (Taber, 2017).

Data Analysis and Discussion

In accordance with standard research conventions, data analysis was conducted following a thorough data cleaning process. This involved excluding incompletely filled questionnaires, addressing missing values through mean replacement, and assessing the normality of the dataset."

Table 1. Demographic profile of respondents

Variable	Frequency (N) (%)	Mean (X)	Standard Deviation (SD)
Gender			
Male	158(69.3)	1.31	0.426
Female	70 (30.7)		
Total:	228(100)		
Age			
18-25 yrs	199 (87.3)	19.24	3.683
26-33 yrs	25(10.96)		
34-41 yrs	3 (1.3)		
>50 yrs	1(0.4)		
Total	228 (100)		
Level			
Year 1	7(3.1)	2.28	0.657
Year 2	168(76.8)		
Year 3	36 (15.7)		
Year 4	16 (7.1)		
Spilling	1(0.4)		
Total	228 (100)		

Table 1 highlights the demographic attributes of the participants (Gender, Age and Level)

Gender: this variable indicate that a majority of participants are male (N=158) while the Female are N=70.

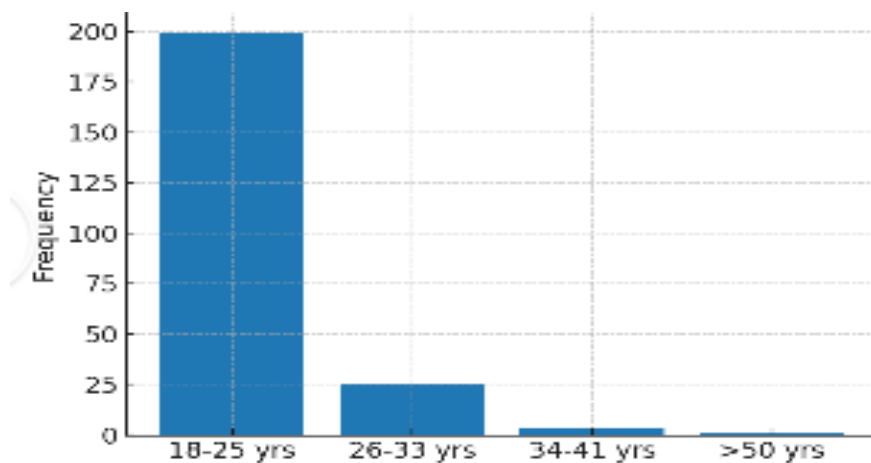


Figure 2 Age distribution of the respondents

From Figure 2, the most populous age of the participant ranges between 18 to 25 years, (N= 199), this is followed by the age group 26 to 33 years (N= 25) reflecting a young participant population, with a minimum of 18 and maximum of 50.

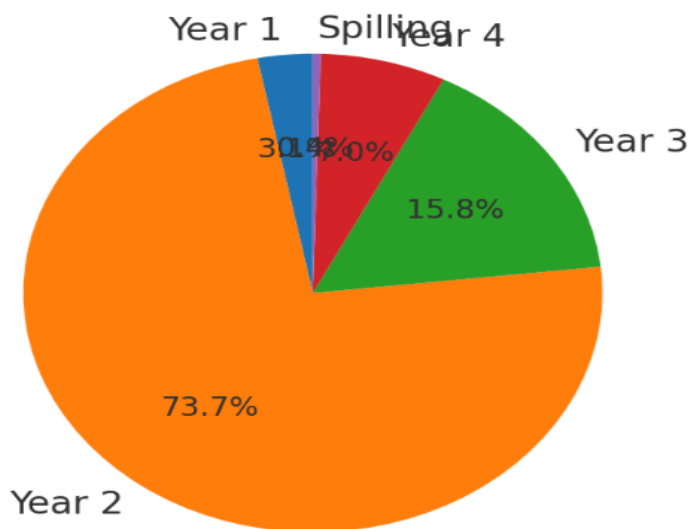


Figure 3: Pie chart showing respondent's distribution by level of study

Figure 3 illustrate the level of study of the computer education students, majority of the participants are in their 2nd year (N= 168, 73.7%), followed by those in the 3rd year

(N=36, 15.7%) while the final (year 4) year students, the least was the year 1 (N=7) perhaps because they are just coming into the system.

The average mean score (M) and standard deviation (SD) in the descriptive analysis is presented in Table 2.

Table 2. Descriptive Statistics of the Constructs Mean & Standard Deviation (M & SD)

Composite Constructs	Mean	Std. Deviation
AI_Awareness	3.2781	1.05449
AI_Perception	3.4956	.99391
AI_Policies	3.2526	.96459
AI_Recommendation	3.4991	1.01452
Valid N (listwise)		

Likert Scale: "1= Strongly Disagree, 5= Strongly Agree"

Table 2 presents the descriptive statistics of the four composite constructs related to students' ethical utilization of AI-Generated Contents derived from a sample of 228 participants. The statistics show the number of valid responses (N), the mean (X) and the standard deviation (SD) for each construct. The results shows that the AI-Institutional Support and Recommendation construct reported the highest mean value (X=3.499, SD=1.015), indicating a stronger agreement among the computer education students for support and recommendation toward ethical use of AIGC. Thus, suggesting the students are open to positive AI initiatives given the right infrastructure and communication. This is followed closely by the AI perception

construct ($M=3.496$, $SD=0.994$). This result also suggests a positive outlook toward ethical use of AIGC in the university. The AI-Awareness and AI Policies constructs had slightly lower mean scores ($M=3.278$, $SD=1.054$; and $M=3.252$, $SD=0.964$) respectively. These indicate moderate levels of awareness of ethical use of AIGC and perception of the ethical use, as well as, policy frameworks regarding AI. In addition, this also underscores the need (due to the gap exposed) for targeted awareness and policy communication in the university. Though, perception and endorsement of AIGC is quite high. It may be a signal that enthusiasm about AI may be growing faster than the infrastructure and regulations needed to support AI's ethical usage. As can be seen the standard deviation numbers across all the constructs are fairly consistent ranging from approximately 0.96 to 1.05, thus reflecting moderate variability in the participant responses.

Recommendations Based on Findings

Based on the findings the following recommendations may suffice:

- i. The university authority and policy makers should consider targeted campaigns, workshops and training to scale up awareness of ethical use of AIGC.
- ii. The management of the institution should leverage the positive AI perception to initiate AI-policies that promote responsible use of AIGC with minimal resistance.
- iii. There is the need to improve clarity, accessibility and enforcement of AIGC-related policies through transparent communication and clear ethical guidelines.
- iv. Investment is required in AIGC implementation strategies, pilot programme based on the positive sentiments of the students towards AIGC.

Limitations of the study

The limitations of this study are as follows:

- a. The study is cross-sectional as such limits the depth of its outcome for generalization.
- b. The constructs are based on self-reported data, which is subject to biases such as social desirability and limited self-awareness
- c. The data does not account for the role of demographic factors such as age, gender, and level of study which might influence perception and awareness differently.

Conclusion

This study examined students' ethical use of AI-generated content in a university in Northern Nigeria through a descriptive survey research approach. The findings reveal generally positive perceptions and support for AI integration, alongside moderate levels of awareness and understanding of AI-related policies. These results suggest a foundational readiness for AI adoption among students, yet underscore the need for targeted educational initiatives and clearer policy communication. While descriptive in nature, the insights offer a meaningful starting point for further inferential and longitudinal research. To ensure ethical and effective AI utilization in academic contexts, institutions must adopt a proactive strategy that enhances awareness, aligns perceptions with policy frameworks, and fosters informed engagement. In closing, it is important to note that as the AI revolution continues to impact education across contexts, it must be guided by contextual awareness on the part of the students and the educators as well as responsible and ethical use of the technology within the ambit of respect for educators role in shaping meaningful learning experiences (Novita, 2025).

Acknowledgments

We wish to acknowledge and appreciate all the students of Computer Science Education for consenting and participating in this study despite the short notice. We also appreciate the support of Dr. Ismaila T. Sanusi at the University of Eastern Finland for his support and encouragement

References

- Chatterjee, S., & Bhattacharjee, K. K. (2020). Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modeling. *Education and Information Technologies*, 25(5), 3443–3463. <https://doi.org/10.1007/s10639-020-10125-9>
- Kasneji, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., & Krusche, S. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Konstantinos, K. T. (2024). Integrating artificial intelligence in science education: Benefits and challenges. *International Journal of Educational Innovation*, 6(3). ISSN 2654-0002.
- Ogunbona, P. (2024, November 19–21). Assessment in the age of AI. Invited keynote address presented at the 9th Faculty of Education and Extension Services (FEES) Annual National Conference, Usmanu Danfodiyo University, Sokoto, Nigeria.
- Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 533–544. <https://doi.org/10.1007/s10488-013-0528-y>
- Sanusi, I. T., Agbo, F. J., Dada, O. A., Yunusa, A. A., Aruleba, K. D., Obaido, G., Olawumi, O., & Oyelere, S. S. (2024). Stakeholders' insights on artificial intelligence education:

Perspectives of teachers, students, and policymakers. *Computers and Education Open*, 7, 100212. <https://doi.org/10.1016/j.caeo.2024.100212>

Sdenka Zobeida, S.-P., Kejiang, X., & Xinyun, H. (2023). Artificial intelligence and learning analytics in teacher education: A systematic review. *Education Sciences*, 12(8), 1–18. <https://doi.org/10.3390/educsci12080569>

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education*, 48(6), 1273–1296. <https://doi.org/10.1007/s11165-016-9602-2>

Uche, A. O., Obiora, A. V., & Nwabudike, F. C. (2025). Ethical considerations and applications of AI-generated content in communication and media studies in Nigeria. *British Journal of Contemporary Education*, 5(1), 48–58. <https://doi.org/10.52589/BJCE-UWEIP6RZ>

Usman, A., Obielodan, O. O., & Yunusa, A. A. (2024). Ethical implications of utilising AI-generated content for academic research in Nigerian institutions. *International Journal of Innovative Technology Integration in Education*, 7(1), 153–162.

Yunusa, A. A., & Yusuf, N. (2024). Harnessing artificial intelligence for enhanced open, distance and flexible education: Prospects and challenges. In *Proceedings of the 44th Annual International Conference of Educational Media and Technology (EMTAN)*, Gombe, Nigeria.

Yunusa, A. A., & Yusuf, N. (2025). Prospects and impediments of the application of artificial intelligence for enhanced open, distance and flexible education. *Rima International Journal of Education*, 4(4). (In press).



National Smart Innovation Conference

IMPACT OF AI TOOLS ON ACADEMIC WRITING AND RESEARCH PRACTICES IN NIGERIAN INSTITUTIONS

Impact of AI Tools on Academic Writing and Research Practices in Nigerian Institutions

Moses, O. D.

Department of Business Management, MIVA Open University, Abuja

Kayodedamola99@gmail.com

Abstract

The study assessed the impact of artificial intelligence (AI) tools on academic writing and research practices among students in Nigerian institutions. The study adopted a cross-sectional survey design. Data were collected from 302 students across universities, polytechnics, and colleges using an online questionnaire (Google Form). Findings revealed that students significantly rely on AI tools for various academic tasks, including grammar and spelling checks, paraphrasing, literature searches, data analysis, and essay writing, with ChatGPT being the most widely used tool. Beyond these regular uses, students reported improvements in writing structure, coherence, research competence, and information synthesis. Moreover, AI tools positively influenced students' cognitive engagement and intrinsic motivation, although this evidence was based on self-reported perceptions. Digital literacy was found to be a critical enabling factor, facilitating effective use of AI tools, while institutional ethical policies influenced students' decisions on AI tools usage; there is a gap in policy communication.

Keywords: *AI tools, academic writing, research practices, ChatGPT.*

Introduction

Just like the earth would be void without light, air, and water, today's world would be incomplete without the internet and digital technologies. The rise of global internet penetration and continuous advancement in digital technologies have facilitated the integration of Artificial Intelligence (AI) in all walks of life (Tas & Gulcu, 2019). AI is a “branch of computer science that focuses on simulating intelligent behaviour in computers and the capacity to mimic, and ideally enhance human behaviour” (Naqvi, 2020). In education, the use of AI tools is rapidly expanding, offering significant opportunities to personalise learning, reduce human error, increase student engagement, simplify complex ideas, support automated assessment, and assist with various administrative and academic tasks such as writing and researching (Duque *et al.*, 2024).

In most institutions in Nigeria, the process of academic writing and research is a complex one (Obateru, 2022). Traditionally, students are often required to spend hours or days in the library, physically scanning shelves to comprehend various methodologies, terminologies, concepts, and manual data analysis alongside their other daily academic workload. It also involved extensive proofreading, formal writing style, logical structure, coherence, originality, and credibility (Gupta *et al.*, 2022; Monu *et al.*, 2020; Huisman *et al.*, 2018). Mastering the formal writing style and specific terminology can be difficult, especially for students undertaking research writing for the first time (Morris, 2018). This traditional method fosters cognitive engagement with research material and methodological rigour, which enables a richer understanding of their field. It embeds critical thinking into every stage of inquiry. The traditional approach, however, does not provide personalised instruction and overlooks differences in individual learning styles and paces, which can lead to

inadequate learning experiences. Its rigidity makes it difficult to address the needs of learners with diverse profiles, and thus, leaves some learners to fall behind while others progress quickly (Hu, 2024; Monu *et al.*, 2020).

In response to these challenges, AI tools have become useful tools, transforming academic tasks such as data analysis, literature reviews, referencing, and collaborative writing (Efebeh *et al.*, 2024). AI tools such as Google Gemini, ChatGPT, Semantic Scholar, and Elicit can assist in carrying out literature reviews, extracting key information from journals and databases, performing data analysis, and visualising tasks, which enable researchers to work with both quantitative and qualitative data (Efebeh *et al.*, 2024; Kung, 2023). Grammarly and Quillbot can help students improve writing quality through paraphrasing and reorganising text without losing meaning and aid with grammar checks and corrections (Latifah *et al.*, 2024; Dewi, 2022). These tools have become not just helpful but increasingly essential for improving both the efficiency and quality of academic writing and research.

Following the Technology Acceptance Model (TAM), students' willingness to adopt AI tools depends on two factors: perceived usefulness and simplicity of use (Davis, 1989). If students feel that AI tools can genuinely improve their academic writing and research and are simple to use, they will incorporate them into their learning processes. Collaboration between AI and humans in academic activities has advantages to both students and society, due to it allows students to devote more attention to the critical and innovative aspects of their studies (Nguyen *et al.*, 2024). From a Self-Determination Theory perspective, AI tools can improve students' competence and proficiency by providing instant feedback, fostering autonomy and relatedness through personalised assistance and collaborative effort with AI tools

(Ryan & Deci, 2017). While academic writing and research can be challenging for students, AI tools can help enhance research productivity.

Despite this enhanced productivity, there is a growing concern about the originality and credibility of the content generated by AI tools, their influence on personal development, the potential of decline of student engagement, the fears that the easy access to get information and automated support may lead to academic laziness, and reduce inner drive to engage in difficult academic tasks which is crucial for academic improvement (Marín *et al.*, 2025; Olarewaju *et al.*, 2025; Vieriu, & Petrea, 2025). AI tools such as Grammarly have been reported to produce abstract word suggestions that may not necessarily correspond with a writer's intention, and constant usage of it might lead to students' inability to internalise grammatical rules (Ummah & Bisriyah, 2022). Kurniati and Fithriani (2022) stated that Quillbot improve students' writing skills, yet, there are concerns that it might lead to passive learning behavior. Okebukola (2022) observed that the convenience provided by AI can lead to a decline in critical thinking and problem-solving abilities, both of which are vital components of self learning. Similarly, Qadir's (2023) study found that there is a possibility of students using AI tools in dishonest or unauthorised ways, such as employing AI tools to complete assignments without proper evaluation.

However, in many developed countries such as the UK, AI tools are progressively integrated into education with clear institutional policies (Okunlola, 2024). Conversely, Nigerian institutions and lecturers often resist students from using AI tools for any academic task (Asogwa, 2024; Kayode & Odumabo, 2024). While some institutions and lecturers view AI as a prospect to advance academic outcomes, others see it as a danger to academic integrity. Initially, many Nigerian institutions outrightly banned AI

as academic misconduct and sanctioned students for using it (Kayode & Odumabo, 2024). The concerns were that AI tools enable plagiarism, pose a danger to scientific output, produce biased or erroneous information, and lack precision for application in teaching and learning, due to their still being in a developmental stage (Olawajuwaju *et al.*, 2025; Asogwa, 2024; Kayode & Odumabo, 2024). Mfon *et al.* (2024) found that only 50% of Ritman University students are familiar with AI tools and use them for academic purposes. In addition, Adigun and Okebukola (2023) observed that poor internet connectivity, outdated computing resources, and recurring power outages limit AI integration in many higher institutions in Nigeria. For example, only 48.81% of Nigeria's population currently has access to the internet (Okamgba, 2025).

Given these complexities, this study aims to investigate how students in Nigerian institutions are using AI tools for academic writing and research practices. By examining both the benefits and limitations of these tools, this study aims to provide insights that can inform institutional policies, improve research culture, and educational outcomes in Nigeria, while aligning with global digital education trends.

Statement of the Problem

In spite of the rapid integration of AI tools in education worldwide, their use for academic writing and research within Nigerian institutions is limited, inconsistent, and often misunderstood. While AI tools can support students with handling various complex academic tasks such as data analysis, academic writing, literature review, proofreading, and referencing, its effectiveness is often impeded by several challenges such as resistance and poor digital infrastructure of Nigerian institutions, low digital literacy, internet access, inconsistent power supply and absence of clear policies regulating the ethical and academic use of AI tools. As a result, both students and

academic staff are left uncertain about what constitutes acceptable practice (Okamgba, 2025; Olarewaju *et al.*, 2025; Asogwa, 2024; Kayode & Odumabo, 2024). These limitations risk broadening the digital divide between students in Nigerian institutions and their global counterparts, who are leveraging AI tools to improve research productivity and academic performance (Okunlola, 2024). For example, in many institutions in the UK and Finland, students are obligated to disclose the degree to which they have used AI tools in their academic work. However, there are general concerns that overreliance on AI may erode students' critical thinking, evaluative skills, and intrinsic motivation, as they become passive recipients of AI-generated content without deep cognitive engagement (Marín *et al.*, 2025; Olarewaju *et al.*, 2025; Vieriu & Petrea, 2025). Taking these factors into account, this study aims to assess how students in Nigerian institutions are currently using AI tools for their academic writing and research practices.

Research Questions

1. To what extent do students rely on AI tools for different aspects of academic writing and research tasks?
2. How does using AI tools affect students' academic writing and research competence?
3. What effect do AI tools have on students' cognitive engagement and intrinsic motivation for academic writing and research?
4. How does students' digital literacy influence their use of AI tools for academic writing and research?
5. How do institutions' ethical policies influence students' use of AI tools for academic writing and research?

Method

In this research, a cross-sectional survey design was adopted due to it allows for immediate data collection from a large respondents and ensures that results reflect broad trends rather than specific individual experiences. The study population comprised all undergraduate and postgraduate students in Nigerian tertiary institutions, estimated at 2,500,000 (Asquare, 2024). Using the Taro Yamane method, a sample size of 400 students was determined, and respondents were randomly chosen to take part in the research. Primary data were gathered via an online questionnaire administered through Google Forms. The Google link was shared on different institutional platforms, such as Telegram and WhatsApp, for students to participate. The survey link was distributed across institutional platforms such as Telegram and WhatsApp to reach students. Collected data were analysed using descriptive statistics, such as mean, standard deviation (SD), and frequency. Decision rule for mean scores were interpreted as Strongly Disagree (SD) = 1.00-1.49, Disagree (D) = 1.50-2.49, Neutral (N) = 2.50-3.49, Agree (A) = 3.50-4.49, and Strongly Agree (SA) = 4.50-5.00.

Results

The sample consisted of 302 respondents (75.5% of the 400 sample size). Out of the total respondents, 180 (59.6%) were undergraduates, 122 (40.4%) were postgraduates from diverse institutions, with the majority being from 35 universities (89.7%), followed by 2 Polytechnics (5.1%) and 2 Colleges (5.1%). Almost all respondents reported owning a personal digital device such as a smartphone or laptop (99.0%) and having constant internet access (99.7%). ChatGPT (100%) was the most widely used AI tool, followed by Grammarly (26.5%), Gemini (23.8%), QuillBot (18.2%), Deepseek (4.0%), Meta AI (3.3%), Claude (2.3%), Semantic Scholar (2.0%), Microsoft

Copilot AI (1.3%), Elicit (1.3%), Grok (1.0%), Pi.ai (1.0%), Perplexity (0.7%), NotebookLM (0.3%), Monica (0.3%), and Ai Poe (0.3%).

Table 1: Students' Reliance on AI Tools for Different Aspects of Academic Writing and Research Tasks

	N	Min	Max	Mean	SD
I use AI tools for grammar, punctuation, and spelling checks in my academic writing	302	1	5	3.96	1.053
I rely on AI tools for paraphrasing and improving the clarity and conciseness of my sentences	302	1	5	3.89	.982
I use AI tools for literature searches and summarising journal articles effectively	302	1	5	3.87	.967
I use AI tools for Data analysis and interpretation	302	1	5	3.73	1.046
I use AI tools to assist in writing parts or full essays and research projects	302	1	5	3.88	.943
				Group Mean =	Group SD =
				3.88	0.998

The result revealed that respondents agreed they rely on AI tools for different aspects of academic writing and research tasks. Specifically, the mean score for using AI tools for grammar, punctuation, and spelling checks was 3.96 (SD = 1.053), while reliance on AI for paraphrasing and improving clarity and conciseness was 3.89 (SD = 0.982). Respondents also agreed they use AI tools for literature searches and summarising journal articles (Mean = 3.87, SD = 0.967), data analysis and interpretation (Mean = 3.73, SD = 1.046), and for writing parts or full essays and research projects (Mean = 3.88, SD = 0.943). The group mean was 3.87, and the group SD was 0.998, indicating

that students rely on AI tools for different aspects of academic writing and research tasks.

Table 2: Effect of AI tools on Students' Academic Writing and Research Competence

	N	Min	Max	Mean	SD
Using AI tools has improved my writing structure and coherence	302	1	5	4.05	.828
AI tools have enhanced my ability to conduct thorough and efficient academic research	302	1	5	4.07	.845
AI tools have helped me produce higher-quality academic research write-ups than I could without them	302	1	5	3.97	.876
AI tools have improved my ability to understand and synthesise complex academic information	302	1	5	4.08	.762
AI tools help me complete my academic writing and research tasks faster while maintaining clarity and quality	302	1	5	4.09	.748
				Group Mean = 4.05	Group SD = 0.81

The results showed that respondents agreed that using AI tools positively affects their academic writing and research competence. Specifically, the mean score for AI improving writing structure and coherence was 4.05 (SD = 0.828). Enhancing their ability to conduct thorough and efficient academic research was a mean score of 4.07 (SD = 0.845). Respondents also agreed that AI tools helped them produce higher quality academic research write-ups with mean score of 3.97 (SD = 0.876), improved their ability to understand and synthesise complex academic information (Mean = 4.08, SD=0.762), and assisted them in completing writing and research tasks faster while maintaining clarity and quality (Mean=4.09, SD = 0.748). The group mean score was 4.05, and the group SD was 0.81, indicating that AI tools significantly enhance students' academic writing and research competence.

Table 3: Effect of AI tools on Students' Cognitive Engagement and Intrinsic Motivation for Academic Writing and Research

	N	Min	Max	Mean	SD
Using AI tools keeps me engaged in my academic writing and research tasks	302	1	5	3.83	.819
AI tools motivate me to complete my research write-up more effectively	302	1	5	3.81	.918
I feel more confident about my research write-up when I use AI tools	302	1	5	3.70	.984
Using AI tools increases my interest in academic research and enhances my writing experience	302	1	5	3.78	.908
I am more likely to explore complex topics due to AI assistance	302	1	5	4.07	.905
				Group Mean	Group SD
				= 3.84	= 0.81

The results revealed that respondents agreed that AI tools positively influence their cognitive engagement and intrinsic motivation for academic writing and research. Specifically, using AI tools to stay engaged in academic tasks had a mean score of 3.83 (SD = 0.819), while AI tools motivating students to complete research write-ups more effectively had a mean of 3.81 (SD = 0.918). Likewise, reported they felt more confident about their research write-ups when using AI tools (Mean = 3.70, SD = 0.984), agreed that AI increased their interest in academic research and enhanced their writing experience (Mean = 3.78, SD = 0.908), and increased their likelihood to explore complex topics due to AI assistance (Mean = 4.07, SD=0.905). The group mean was 3.84, and the group SD was 0.91, indicating that AI tools have a positive effect on students' cognitive engagement and intrinsic motivation in their academic writing and research tasks.

Table 4: Effect of Students' Digital Literacy on the Use of AI Tools for Academic Writing and Research

	N	Min	Max	Mean	SD
My proficiency with digital technologies enables me to effectively utilise AI tools for academic writing and research tasks	302	1	5	4.13	.782
My confidence in using digital tools makes it easier for me to adapt to and use AI tools for academic writing and research tasks	302	1	5	4.04	.730
I can easily search for and access AI tools online	302	1	5	4.26	.722
I can evaluate the outputs generated by AI tools critically	302	1	5	4.05	.822
b	302	1	5	4.19	.827
				Group Mean = 4.13	Group SD = 0.78

The results showed that respondents agreed that their digital literacy influences their use of AI tools for academic writing and research. Specifically, proficiency with digital technologies enabling effective use of AI tools had a mean score of 4.13 (SD = 0.782), and confidence in using digital tools to adapt to AI tools had a mean score of 4.04 (SD = 0.730). Students also agreed they can easily search for and access AI tools online (Mean = 4.26, SD = 0.722), critically evaluate outputs generated by AI tools (Mean = 4.05, SD = 0.822), and that their knowledge of digital technologies helps them maximise AI tools (Mean = 4.19, SD=0.827). The group mean was 4.13 and the group SD was 0.78, indicating that students' digital literacy positively influences their effective utilisation of AI tools for academic writing and research tasks.

Table 5: Impact of Institutions' Ethical Policies on Students' Use of AI Tools for Academic Writing and Research

	N	Min	Max	Mean	SD
My institution has clear ethical policies on AI use for academic work	302	1	5	3.89	.978
My institution's ethical policies regarding AI tools are consistently communicated to students	302	1	5	3.68	1.012
I understand the ethical guidelines set by my institution regarding the use of AI tools in academic work	302	1	5	3.86	.906
I am aware of the specific consequences for violating my institution's AI usage policies	302	1	5	3.82	.957
The ethical policies at my institution shape my decisions on when and how to use AI tools for academic writing and research	302	1	5	3.88	.976
				Group Mean =	Group SD =
				3.83	0.97

The research results revealed that respondents agreed that institutional ethical policies influence their use of AI tools for academic purposes. Specifically, respondents agreed that their institution has clear ethical policies on AI use (Mean = 3.89, SD = 0.978) and that they understand the ethical guidelines set by their institution regarding AI tools (Mean = 3.86, SD = 0.906). The mean score for consistent communication of these policies was 3.68 (SD = 1.012). Awareness of specific consequences for violating AI usage policies had a mean of 3.82 (SD = 0.957), while the perception that ethical policies shape their decisions on AI use scored 3.88 (SD = 0.976). The group mean was 3.83, and the group SD was 0.97. The results suggest that students generally agree that play an important role in shaping their decisions and behaviours when using AI tools for academic writing and research.

Discussion of Findings

The study results showed that most students were in agreement regarding the use of AI tools for different aspects of academic writing and research, with a group mean of 3.88. The most common uses include essay writing, grammar checks, paraphrasing, literature searches, and data analysis. This aligns with the study by Ngonso et al. (2025), which revealed that the majority of Nigerian tertiary students utilise AI, and it enhances their academic performance. The findings revealed that students were of the opinion that AI tools improve their academic writing and research ability, with a group mean of 4.05. Respondents highlighted that AI tools not only support surface-level tasks such as grammar checking but also enhance more in-depth writing and research skills, such as synthesis of information, writing structure, and the ability to articulate complex research concepts and methodologies. This is relevant in Nigeria's educational system, where students often face limited quality research supervision, poor digital infrastructure and library resources, and poor research training. This also showed that AI tools are not just supplementary tools for students; they directly contribute to mitigating the structural barriers of the traditional academic process, such as limited access to physical resources and the labour-intensive nature of research (Monu et al., 2020). This finding is consistent with Kim *et al.*'s (2025) results, who reported that students expect AI tools to perform multiple functions, such as a virtual tutor, a versatile writing assistant, and even a digital partner that supports the writing process and overall performance. In terms of cognitive engagement and intrinsic motivation, students acknowledged that AI tools keep them engaged and motivated to complete their academic work efficiently, with a group mean of 3.84. This finding is based on self-reported perceptions, which may not fully capture deep cognitive processing. However, from a Self-Determination Theory perspective (Ryan & Deci, 2017), the data suggest that AI tools may satisfy students' basic psychological needs for competence (by providing instant feedback and helping

them produce higher-quality work) and autonomy (by giving them more control over their academic tasks). This satisfaction of needs likely contributes to the reported positive effect on perceived engagement and intrinsic motivation. This aligns with Hanshaw's (2024) study, which revealed that AI improved students' grades and intrinsic motivation significantly. The findings also revealed that the digital literacy of students influences their use of AI tools, with a group mean of 4.13. Competency in digital technologies enabled effective searching, critical evaluation, and optimal use of AI tools for academic purposes. This suggests that without strong digital literacy skills among Nigerian students, the benefits of AI in academic writing and research would be significantly limited, where infrastructural challenges already disadvantage some learners. This finding aligns with Joseph *et al.*'s (2024) study, which revealed that students' digital literacy positively correlated with AI-assisted learning. Finally, the study findings revealed that institutional ethical policies influence students' use of AI tools, with a group mean of 3.83. Students acknowledged that ethical guidelines and policy clarity shape their decisions on how to use AI. However, the mean score is lower in constant communication of these policies, which points to a gap in policy dissemination. This inconsistency suggests that while students are generally aware of AI-related ethical guidelines, they may lack clear and consistent instruction on the boundaries of appropriate use. As a result, the use of AI is more likely to be based on individual interpretation as opposed to institutional guidance. This finding is consistent with a study by Izevbigie *et al.* (2025), which revealed that awareness of the challenges related to AI is the most important predictor of ethical considerations. Familiarity with AI and perceived opportunities did not have a significant impact. The result underscores the need for Nigerian institutions to establish clear ethical frameworks, improve policy communication, increase awareness of potential risks, and promote responsible practices in AI use.

Conclusion

This research assessed the impact of AI tools on academic writing and research practices in Nigerian institutions. Based on the findings, the study concluded that students use AI tools for grammar checks, paraphrasing, literature searches, data analysis, and essay writing. Beyond these surface-level uses, it was concluded that AI tools enhance writing structure, coherence, efficiency, and information synthesis. However, these perceived benefits are particularly relevant within Nigeria's educational system, where students face challenges such as limited quality supervision, poor digital infrastructure, and insufficient library resources. Thus, AI tools function not merely as supplementary tools but as a direct means of addressing structural barriers such as limited access to resources and the labour-intensive nature of research. Likewise, AI tools influenced students' cognitive engagement and intrinsic motivation positively, although this evidence is based on self-reported perceptions and may not fully capture deeper cognitive processes. Digital literacy was identified as a strong enabling factor for effective use of AI tools, while institutional ethical policies shaped students' decisions on the use of AI tools. However, there is a gap in policy communication. Therefore, Nigerian institutions should establish and provide clear communication on ethical frameworks guiding the use of AI tools for all academic activities to promote responsible integration of AI without compromising academic integrity. However, this study is limited by the reliance on self-reported data, which means that the results reflect students' perceptions rather than direct measures of learning outcomes or cognitive engagement. The cross-sectional design limits the ability to establish causal relationships between AI use and improvements in academic writing or motivation. Likewise, primary reliance on quantitative data did not capture the depth and complexity of students' experiences with AI. Future research should

therefore adopt mixed-methods approaches, which will allow researchers to capture not only the prevalence and patterns of AI use but also the complex experiences, motivations, and challenges when using AI tools. Longitudinal or experimental designs could be adopted to examine the long-term effects of AI on cognitive engagement. Further studies could also investigate lecturers' and supervisors' perspectives on AI use, and compare AI adoption trends across different cultural institutions.

References

- Adigun, H. O., & Okebukola, P. A. (2023). Leveraging artificial intelligence for enhanced assessment and feedback mechanisms in Nigeria's higher education system. *International Journal of Research in Innovative Science and Society*, 8(9), 142-151.
- Asogwa, V. C. (2024). Existing and future challenges faced by Nigerian university teachers in the artificial intelligence era. *Journal of Science, Technology, Education, and Management*.
- Asquare, A. (2024). How many students are in Nigerian tertiary institutions? Students Media.
<https://studentsmedia.com.ng/howmany-students-are-in-nigeria-tertiary-institutions/>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
<https://doi.org/10.2307/249008>
- Dewi, U. U. (2022). Grammarly as automated writing evaluation: Its effectiveness from EFL students' perceptions. *Lingua Cultura*, 16(2), 155-161.
<https://doi.org/10.21512/lc.v16i2.8315>
- Duque, R. D. C. S., Placido, R. L., Monteiro, E. L., Monteiro, R. R., Fraga, R. P., De Oliveira, R. N., Paulino, L. D. S., Assunção, I. D., Da Silva, J. J., & Da Costa, S. D. S. R. (2024). Artificial

intelligence and teaching practice: Concepts, applications and educational challenges. *IOSR Journal of Business and Management*, 26(10), 24-30. <https://doi.org/10.9790/487X-2610092430>

Efebeh, V. E., Orishede, F., & Igoh, J. M. (2024). Artificial intelligence and academic research in contemporary society: Evidence from university academics. *Ianna Journal of Interdisciplinary Studies*, 6(3), 33-44. <https://doi.org/10.5281/zenodo.13152543>

Gupta, S., Jaiswal, A., Paramasivam, A., & Kotecha, J. (2022). Academic writing challenges and supports: Perspectives of international doctoral students and their supervisors. *Frontiers in Education*, 7, Article 891534. <https://doi.org/10.3389/feduc.2022.891534>

Hanshaw, G., Vance, J., & Brewer, C. (2024). Exploring the effectiveness of AI course assistants on the student learning experience. *Open Praxis*, 16(4), 627-644. <https://doi.org/10.55982/openpraxis.16.4.719>

Hu, J. (2024). The challenge of traditional teaching approach: A study on the path to improve classroom teaching effectiveness based on secondary school students' psychology. In *Proceedings of the 2nd International Conference on Social Psychology and Humanity Studies* (pp. 945-950). <https://doi.org/10.54254/2753-7048/50/20240945>

Huisman, B., Saab, N., van Driel, J., & van den Broek, P. (2018). Peer feedback on academic writing: Undergraduate students' peer feedback role, peer feedback perceptions and essay performance. *Assessment & Evaluation in Higher Education*, 43(6), 955-968. <https://doi.org/10.1080/02602938.2018.1424318>

Izevbogie, H. I., Olajide, O., Olaniran, O., & Akintayo, T. A. (2025). The ethical use of generative AI in the Nigerian higher education sector. *World Journal of Advanced Research and Reviews*, 25(3), 847-853. <https://doi.org/10.30574/wjarr.2025.25.3.0789>

- Joseph, G. V., Athira, P., Thomas, M. A., Jose, D., Roy, T. V., & Prasad, M. (2024). Impact of digital literacy, use of AI tools and peer collaboration on AI-assisted learning: Perceptions of university students. *DER*, 45, 43-49. <https://doi.org/10.1344/der.2024.45.43-49>
- Kung, J. Y. (2023). Elicit. *The Journal of the Canadian Health Libraries Association*, 44(1), 15-18. <https://doi.org/10.29173/jchla29657>
- Kurniati, E. Y., & Fithriani, R. (2022). Postgraduate students' perceptions of Quillbot utilization in English academic writing class. *Journal of English Language Teaching and Linguistics*, 7(3), 437-451. <https://doi.org/10.21462/jeltl.v7i3.852>
- Latifah, S., Muth'im, A., & Nasrullah, N. (2024). The use of QuillBot in academic writing: A systematic literature review. *Journey Journal of English Language and Pedagogy*, 7(1), 110-121. <https://doi.org/10.33503/journey.v7i1.872>
- Marín, R., Caro, O. C., Rubio, Y. d. C. M., Tuesta, J. N. A., Bardales, E. S., Rituay, A. M. C., & Santos, R. C. (2025). Artificial intelligence as a teaching tool in university education. *Frontiers in Education*, 10, Article 1578451. <https://doi.org/10.3389/feduc.2025.1578451>
- Monu, J. O., Bamgbose, A. A., & Okunnu, H. O. (2020). Effective use of library and its impact on students' study habits in selected universities in Lagos State, Nigeria. *Library and Information Science Digest*, 13, Article 6, 74-84.
- Morris, E. J. (2018). Academic integrity matters: Five considerations for addressing contract cheating. *International Journal for Educational Integrity*, 14(1), 15. <https://doi.org/10.1007/s40979-018-0030-0>
- Naqvi, A. (2020). *Artificial intelligence for audit, forensic accounting, and valuation: A strategic perspective*. Wiley.
- Ngonso, B. F., Egielewa, P. E., Egenti, G., Uduehi, I., Sunny-Duke, F., Ukhurebor, K. E., Onwusinkwue, S., Odezuligbo, I., Abiodun, A. O., Talabi, A. A., Jokthan, G., Opataye, J., Nwankwo, U. C., Eneche, B. M., & Osemengbe, U. O. (2025). Influence of artificial

intelligence on educational performance of Nigerian students in tertiary institutions in Nigeria. *Journal of Infrastructure, Policy and Development*, 9(1), Article 9949. <https://doi.org/10.24294/jipd9949>

Nguyen, A., Hong, Y., Dang, B., & Huang, X. (2024). Human-AI collaboration patterns in AI-assisted academic writing. *Studies in Higher Education*. Advance online publication. <https://doi.org/10.1080/03075079.2024.2323593>

Obateru, O. T. (2022). Research projects and dissertation writing among Nigeria university students: Perceived difficulty areas and the way out. *Sapientia Foundation Journal of Education, Sciences and Gender Studies*, 4(1), 141-150.

Okamgba, J. (2025, July 15). Internet penetration grows 21% in five years. Punch Newspaper. <https://punchng.com/internet-penetration-grows-21-in-five-years/>

Okunlola, J. O. (2024). Investigating perceptions and practices of plagiarism among Nigerian undergraduates in the fourth industrial revolution era: The role of ChatGPT. *E-Journal of Humanities, Art and Social Sciences*, 1796-1808.

Olarewaju, Y., Hammed, L. A., & Raheem, A. O. (2025). Impact of artificial intelligence tools on student initiative: Examining academic laziness among education undergraduate students at University of Ilorin. *Journal of Contemporary Education Research*, 7, 147-158. <https://doi.org/10.70382/hujcer.v7i8.012>

Qadir, J. (2023, May 1-4). Engineering education in the era of ChatGPT: Promise and pitfalls of generative AI for education. In 2023 IEEE Global Engineering Education Conference (EDUCON) (pp. 1-9). IEEE. <https://doi.org/10.1109/EDUCON55603.2023.10151572>

Ryan, R. M., & Deci, E. L. (2017). *Self-determination theory: Basic psychological needs in motivation, development, and wellness*. Guilford Press.

Taş, N., & Gülcü, A. (2019). *Bilgisayar destekli öğretimin temelleri [Fundamentals of computer aided instruction]*. Erzurum: Ertual Akademi.

Ummah, L. K., & Bisriyah, M. (2022). EFL students' perception on Grammarly's premium's feedback and dealing with inaccuracies. *Journal of English Educators Society*, 7(2), 163-172. <https://doi.org/10.21070/jees.v7i2.1687>

Vieriu, A. M., & Petrea, G. (2025). The impact of artificial intelligence (AI) on students' academic development. *Education Sciences*, 15(3), 343. <https://doi.org/10.3390/educsci15030343>

ARTIFICIAL INTELLIGENCE ENABLED RANSOMWARE: A SYSTEMATIC REVIEW

Artificial Intelligence Enabled Ransomware: A Systematic Review

¹Osin, O. J., ²Isah A.O. ³Subairu, S.O., ⁴Suleiman, A., & ⁵Noel, M.D.

¹Federal University of Technology, Minna, ²Federal University of Technology, Minna,

³Federal University of Technology, Minna, ⁴Federal University of Technology, Minna

& ⁵Federal University of Technology, Minna

Corresponding author: osin.oluwatosin@st.futminna.edu.ng

Abstract

Ransomware has been one of the most severe cyberthreats, hobbled operations as well as bottom lines worldwide. Normally, ransomware attacks have been based on the idea of encrypting or blocking the data itself and paying a ransom for it to be released, yet the introduction of AI now also leaves many traditional detection methods out of date. This survey aims to answer three main research questions about how AI-empowered ransomware are growing as a threat: What are these emerging threats, state-of-the-art detection methodologies to detect them and what are the related forthcoming research directions? Applying the PRISMA 2020 methodology, the retrieved papers between 2020 and 2025 were systematically reviewed, where only the ones of high quality and focus were included. What's clear is that AI has turned ransomware from something that was statically built to something that's now intelligent, work rounding, complex. However, there is hope on the horizon for detecting and defending against such attacks due to the advances in AI-based detection and response systems. In the following, we reflect on remaining challenges identified in the study and the importance of further work to bridge the gap and make AI-based defenses more impactful.

Keywords: Artificial Intelligence, Ransomware Detection, Machine Learning, Deep Learning, threat.

1.0 Introduction

Ransomware, a type of harmful cyberattack, has grown more sophisticated thanks to advancements in Artificial Intelligence (AI). Attackers now use AI to automate their methods, adapt quickly, and slip past traditional security systems. By leveraging techniques like machine learning, deep learning, and generative AI, ransomware can improve its strategies, making it more difficult to spot and stop. Recent studies have shown that AI-powered ransomware variants tend to cause more damage (Ferdous et al., 2024; Khan, 2023; Vasan et al., 2020; Alazab et al., 2024; Sharma et al., 2024; Verma, Zeng & Faridi, 2020). This review focuses on summarizing research from 2020 to 2025 that explores how AI is being used in both attacking and defending against ransomware. It looks at key trends, the datasets and criteria researchers use, and points out areas where more work is needed to strengthen cybersecurity defenses against AI-enhanced ransomware. Additionally, the review compares these findings with existing studies to provide a well-rounded understanding of the topic, as outlined in Table 1.

Table 1: Comparison with existing literature

Literature	Offensive AI	Defensive AI	Case Studies	Methodology	Unique Contribution
Ferdous et al., 2024	Limited	✓	X	Narrative Review	Focus on ML/DL Detection
Jasler, 2023	X	✓	X	SLR	Health Focus
Priya et al., 2024	X	✓	✓	Malware Analysis	Technical breakdown
This Work	✓	✓	✓	PRISMA 2020, Multi-database	Dual focus (Offense/Defense, Latest cases & Research Gaps)

2.0 Related Work

This review highlights how AI is making ransomware smarter and harder to detect by improving its evasion tactics. Ferdous et al. (2024) point out challenges in detecting diverse ransomware and suggest advanced AI models that adapt better while protecting privacy. Jasler (2023) notes rising ransomware risks in healthcare, worsened by COVID-19, with AI and blockchain offering new defense options. Priya et al. (2024) study the WannaCry attack, stressing the need for updates, backups, and monitoring to prevent such threats. Jia et al. (2024) discuss machine learning's role in industrial cybersecurity but warn of challenges like adversarial attacks and call for combining AI with traditional methods. Santosh, Govind, and Sanjay (2024) review AI techniques for detecting Android malware, highlighting the need for adaptive systems and fewer false alarms.

3.0 Methodology

This review follows the PRISMA 2020 guidelines, ensuring a transparent and systematic approach. The process includes planning the study, defining clear research questions, carefully selecting and screening relevant papers, extracting key information, and synthesizing the findings. The inclusion and exclusion criteria, detailed in Table 3, were used to choose only those studies that aligned closely with the review's objectives.

3.1 Review Protocol and Research Questions.

This systematic review followed a carefully designed plan based on guidelines from Page et al. (2021). The plan clearly outlined the scope, objectives, and steps of the review.

The main questions it sought to answer were:

- a. What risks does AI-powered ransomware pose?
- b. What methods are currently used to detect AI-based ransomware?
- c. What future directions should research take regarding AI-enabled ransomware?

3.2 Information Source and Search Strategy

This systematic literature review analyzed 1,297 publications on AI-enabled ransomware from various academic databases as shown in table 2 below. After eliminating duplicate entries and carefully applying the selection criteria, a final collection of unique studies was assembled, with the majority sourced from

ScienceDirect. These selected studies serve as the foundation for analyzing current trends and uncovering key themes and gaps in the field.

Searching Strings

The search strategy utilized Boolean operators combining key terms such as (“artificial intelligence” OR “machine learning” OR “deep learning”) AND (“ransomware” OR “AI-enabled ransomware”) AND (“detection” OR “classification” OR “analysis”). To ensure relevance and quality, the search was limited to peer-reviewed articles published between 2020 and 2025, written in English. This approach helped focus on recent advances in AI-driven ransomware research, including detection techniques and behavioral analysis.

Table 2: Database Sources

Database	Rationale
Scopus	Scopus is a crucial multidisciplinary resource for systematic literature reviews, offering a comprehensive collection of abstracts, citations, data, and scholarly content, ensuring scholarly and validated material, and simplifying research trend tracking and topic identification.
IEEE Xplore	IEEE Xplore is a valuable resource for systematic literature reviews due to its high-coverage of engineering records, making it suitable for evidence synthesis in engineering topics, aligning with systematic review guidance.
ScienceDirect	ScienceDirect is a valuable database for systematic literature reviews, offering a vast collection of peer-reviewed, full-text scientific, technical, and health literature across various

ACM Digital Library

subject areas, facilitating comprehensive searches and unbiased evidence synthesis.

The ACM Digital Library is a valuable tool for systematic literature reviews in computer science, offering a focused collection of high-quality publications, advanced search functions, and comprehensive indexing.

Table 3: Inclusion and Exclusion Criteria

Criterion	Inclusion	Exclusion
Timeline	Only studies published within the last 6 (2020-2025) years were included in order to have up-to-date research on the topic	Studies published before 2020
Focus	Studies focused on AI-enabled ransomware or machine learning in ransomware	Studies discussing ransomware generally without AI-enabled focus
Literature Type	Research articles, journals and trusted websites that provides valuable insights on the topic	Review papers, book chapters, series, opinion pieces
Study Type	Research that included empirical analysis, case studies, or surveys	Theoretical papers lacking application or evaluation
Publication Stage	Final	Other than final
Language	English	Non-English
Accessibility	Full-text articles that are readily accessible through the search databases	Those whose full-text is not readily available

3.3 Eligibility Criteria

This review focuses on AI-enabled ransomware analysis and detection, excluding editorials, opinion pieces, and research lacking clear methodology. Studies from 2020-2025 are prioritized, with English accessibility, clear definitions, and full-text availability for thorough assessment.

3.4 Study Selection

The study selection followed the PRISMA 2020 four-step process: identification, screening, eligibility, and inclusion (Page et al., 2021). After removing duplicates, titles and abstracts were screened for relevance, followed by a full-text review. Ultimately, 41 articles met the criteria and were included in the final analysis as shown in figure 1 below.

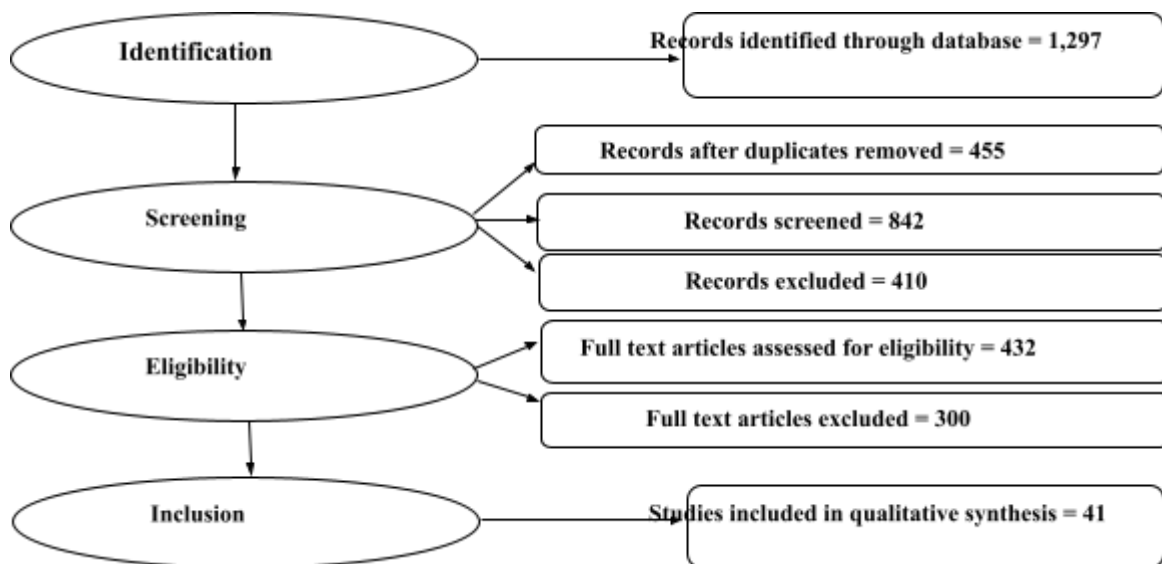


Figure 1. PRISMA Methodology Flow Chart

4.0 Result

This review reveals the global nature distribution of AI-enabled ransomware research, with North America and Europe leading in both offensive and defensive AI ransomware studies, followed by growing contributions from Asia as shown in Figure 2 below.



Figure 2: Research Global Distribution on AI-enabled Ransomware

4.1 Result on Threats Impact of AI-enabled Ransomware

The study reveals a rise in AI-enabled ransomware threats from 2020 to 2025, driven by professionalization of ransomware through AI powered Ransomware-as-a-service platforms, remote work, and increased attack surfaces. Attacks have shifted from mass campaigns to targeted operations targeting critical sectors, often employing double extortion tactics. AI's role in ransomware development has grown, but challenges remain, highlighting the evolving nature of these threats.

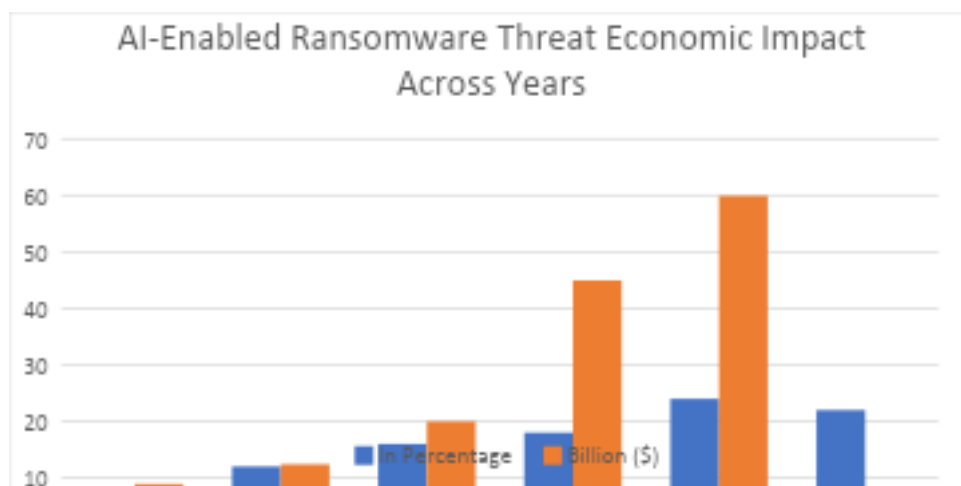


Figure 3: AI-enabled Ransomware Distribution

The figure 3 above reveals that AI-enabled ransomware has significantly impacted the global economy, with the highest economic loss occurring in 2024, resulting in a global loss of \$60 billion. The severity of these attacks, particularly Ascension (Black Basta) shown in figure 4 below, is also significant.

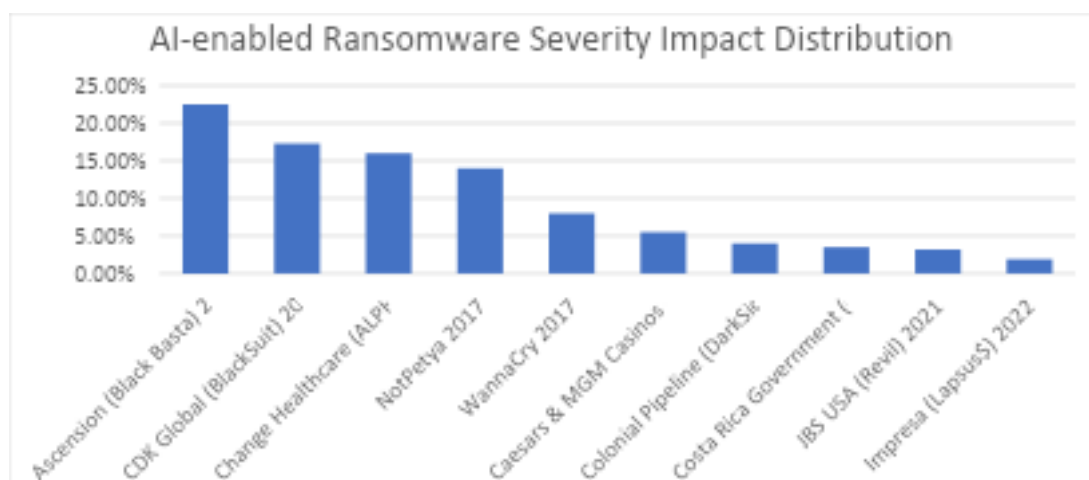


Figure 4: AI-enabled ransomware impact on economy distribution over the years.

4.1.1 Real-World Impact of AI-Enabled Ransomware (Recent Case Studies)

1. BlackCat/ALPHV (2023–2024)

BlackCat, also known as ALPHV, uses AI to extract sensitive data from compromised networks, using phishing campaigns and machine learning algorithms. Their sophisticated approach sets a new standard for ransomware

sophistication, necessitating rethinking of security measures. (Ferdous et al., 2024, Patel (2024)).

2. Royal Ransomware (2024)

Royal ransomware attacks in 2024 targeted financial institutions and municipalities, using machine learning models to mimic normal system processes. These attacks highlighted the need for behavioral analytics and anomaly detection. (Halcyon, 2025; Ferdous et al., 2024).

3. RansomHouse (2023–2025)

RansomHouse, active from 2023–2025, exploited vulnerabilities in enterprise networks using AI-driven reconnaissance tools. In 2025, they targeted a global logistics company, automating cloud service exploitation and coordinating extortion campaigns. This highlights the urgent need for AI-based threat hunting and automated incident response. (Halcyon, 2025; Ferdous, Alazab & Venkatraman, 2024 Sasi et al. (2024)).

4.1.2 AI-enabled Ransomware Analysis across Industries: The evolution of ransomware, which is now even supported by artificial intelligence (AI), has had far-reaching consequences across various industries. Table 4 below present the comparative analysis of AI-enabled ransomware impacts across different industries.

Industries	AI-enabled Tactics	Impact	Challenges
Healthcare Kim & Lee (2022) Yu, Shvetsov & Alsamhi (2024) Czeczot et al. (2023) Wang et al. (2021) Taheri, Kadir & Lashkari	AI attackers use phishing emails to target hospital systems, increasing compromise risk. Machine learning algorithms make ransomware adaptation to network defenses, making detection harder.	Healthcare, heavily reliant on digital records and networked medical devices, has become a prime target for AI-driven ransomware due to the COVID-19 pandemic's digital transformation.	The disruption of patient care and emergency services, financial losses and reputational damage, and the forced return to manual operations during outages are all potential consequences.
Finance Gaber, Ahmed & Janicke (2024) Smmarwar, Gupta & Kumar (2024) Admass, Munaye & Diro (2024)	AI tools automate asset identification, mimic legitimate transactions, and identify outdated or vulnerable systems for exploitation, evading fraud detection systems and enabling attackers to exploit them.	Financial institutions are constantly threatened by AI-enabled ransomware, which often targets payment systems, customer databases, and transaction records due to the high value of their data and assets.	The threat of cyber threats could lead to disruptions in banking operations, potential regulatory penalties, loss of customer trust, and an increased risk of double extortion.
Education Ucci, Aniello & Baldoni (2020) Anandhi et al. (2023)	AI creates spear-phishing emails, automated tools scan for unpatched systems, and ransomware adapts to various IT environments, targeting students and staff.	Universities and schools have become frequent ransomware targets, especially with the rise of remote learning platforms. AI-enabled Tactics:	The threat of cyberattacks could disrupt academic and research activities, potentially leading to the loss of sensitive data and putting financial strain on institutions with limited cybersecurity budgets.
Manufacturing & Logistics Panchal, Khadse &	AI automates supply chain network mapping, identifies production nodes,	Manufacturers and logistics companies are increasingly targeted due to	The disruptions caused by the pandemic could result in

Mahalle; Galli et al. Qamar, Karim & Chang Gihavo et al. Carrier et., al (2022)	exploits IoT vulnerabilities, and uses machine learning models to help malware remain undetected in complex, hybrid environments.	their reliance on industrial control systems and just-in-time supply chains.	significant financial losses, long-term reputational damage, and business loss.	4.2
Government Vinayakumar et al. Verma et. Al (2023) Moti et al. (2021) Canadian Institute for Cybersecurity (2023)	AI scans public systems for vulnerabilities, automates initial exploitation, and enables ransomware to adapt to government networks, prioritizing encryption of critical infrastructure and citizen services.	Government agencies, from local municipalities to national departments, are attractive targets due to the sensitive data they hold and the potential for widespread disruption.	The disruption of essential public services, exposure of confidential citizen data, and erosion of public trust and political fallout are potential consequences.	

Result Evasion Techniques

AI-enabled ransomware employs advanced evasion strategies and adaptive behaviors to enhance its effectiveness and evade detection. These tactics conceal its presence, hinder investigation, and confuse security technologies, making it resistant to both conventional and advanced security defenses (Tripwire, 2024; Halcyon.ai, 2025).

AI-Powered Social engineering has the highest success rate as seen in figure 4 below.

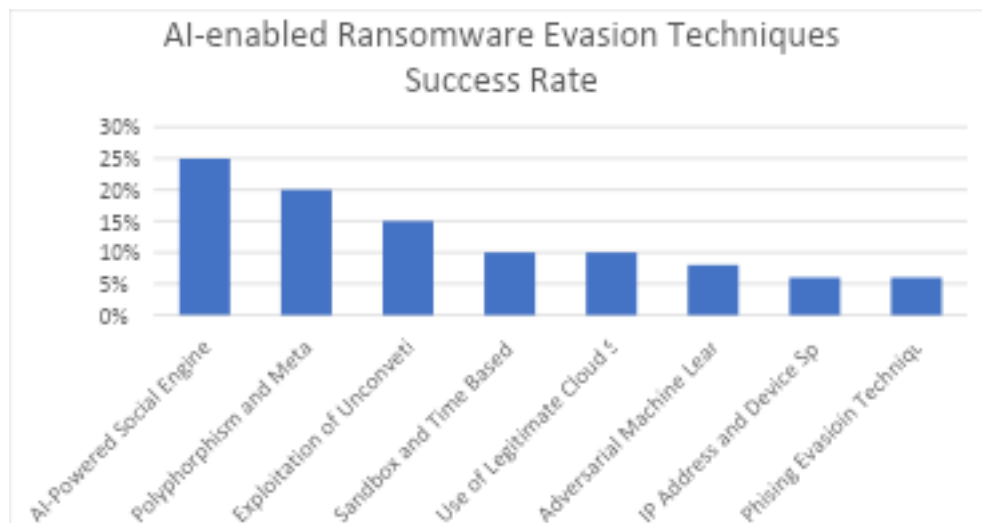


Figure 5: AI-enabled ransomware evasion techniques to avoid detection

4.3 Result on Detection Techniques

From 2020 to 2025, AI-enabled ransomware detection techniques have made significant progress, with machine learning and deep learning models identifying complex patterns and code anomalies as observed from figure 6 below. However, challenges like dataset scarcity, vulnerability to attacks, and scalability remain. Future research should focus on hybrid models, transfer learning, and robust evaluation frameworks to keep pace with advanced ransomware tactics.

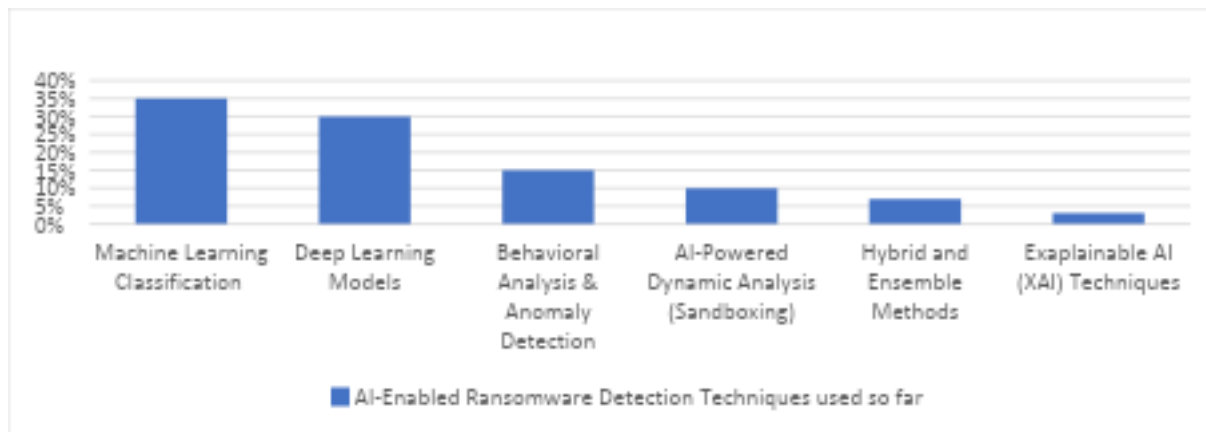


Figure 6: Several AI-enabled Ransomware Detections Techniques been researched till date

4.3.1 Dataset used

This review gives the various dataset being used by researchers in their research as pertaining to AI-enabled ransomware most especially in the detection techniques.

Below are the most recent ones being used: Ransomware Dataset 2024 and UGRansome2024, Dynamic Feature Dataset, MCC and MCFP.

However, these datasets can be combined to address limitations (e.g., MCC's generality + UGRansome2024's network insights).

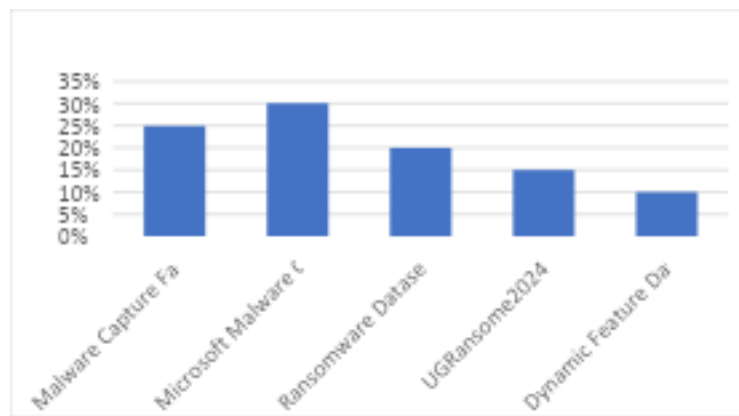


Figure 7: Mostly Used Research Dataset for AI-enabled Ransomware

4.3.2 AI-enabled Ransomware Analysis Techniques

Recent studies since 2020 have improved AI-driven techniques for detecting, classifying, and combating ransomware threats, utilizing machine learning, deep learning, and advanced optimization for enhanced accuracy and reliability.

- 1. Machine Learning-Based Detection:** AI ransomware detection uses supervised, unsupervised, and reinforcement learning methods to identify ransomware and safe datasets. Supervised models focus on relevant data, improving performance through feature engineering and metrics like precision, recall, and F1-score (Ferdous, Alazab, & Venkatraman, 2024).

2. **Deep learning and hybrid approaches:** including recurrent neural networks and LSTM networks, are utilized to understand ransomware behavior patterns and detect evolving tactics (Kumar et al., 2025).
3. Unsupervised learning models like autoencoders and clustering detect unusual system or network behaviors, indicating ransomware activity. Monitoring system calls and file access patterns helps catch ransomware early (Ferdous et al., 2024).
4. Transfer learning and Explainable AI techniques improve adaptability and speed up training for ransomware detection (Ferdous et al., 2024).
5. Lightweight AI models are designed for real-time detection of ransomware, balancing speed and accuracy, making them ideal for resource-constrained environments like IoT devices (Kumar et al., 2025).

4.3.3 Resources/Tools Used in AI-enabled Ransomware

In order to carryout AI-enabled ransomware, there are tools and resources involved, the key resource collated from this review are seen below but might not be limited to these:

1. **Machine Learning and Deep Learning Frameworks:** The frameworks used enable the development of adaptive ransomware that can evade traditional detection methods through feature extraction, classification, and anomaly detection (Hussain et al., 2024).
2. **Automated Feature Engineering and Data Preprocessing Tools:** Essential for extracting meaningful indicators from system logs, memory, and network traffic to enable AI models to identify encryption activity and process anomalies (Hussain et al., 2024).

- 3. Natural Language Processing (NLP) and large language models:** are used to search for important documents, create believable phishing emails, and automate social engineering techniques that help spread ransomware (SoteroSoft, 2024).
- 4. Advanced Encryption Algorithms and Cryptographic Libraries:** AI can dynamically select and apply encryption strategies to maximize damage and complicate decryption efforts by victims or defenders (Alhassan & Alhassan, 2021).
- 5. Automated Reconnaissance and Vulnerability Scanning Tools:** AI-driven scanning tools autonomously identify exploitable system weaknesses and network misconfigurations to facilitate ransomware infiltration (NAKIVO, 2025).
- 6. Behavioral Analysis and Evasion Techniques:** AI models enable ransomware to mimic legitimate system processes and adapt behavior in real-time (Hussain et al., 2024).
- 7. Automation Frameworks for Propagation and Persistence:** Automated tools enable lateral movement, privilege escalation, and persistence mechanisms to maintain control over compromised environments without manual intervention (Patel & Kumar, 2024).

5.0 Discussion of Findings

The integration of AI into ransomware has significantly changed the scale and complexity of cyberattacks, making them more adaptive, elusive, and damaging. The rapid professionalization of ransomware operations, particularly through Ransomware-as-a-Service (RaaS) platforms, has led to an increase in ransomware incidents worldwide. The pandemic has had a particularly severe impact on essential sectors like healthcare, where traditional security approaches have often fallen short.

While AI-driven methods for detecting and countering threats offer encouraging possibilities, challenges remain—such as limited access to quality data, the need for AI systems that provide clear explanations for their decisions, and the risk that these models could be tricked or manipulated.

5.1 Conclusion

AI-powered ransomware is becoming an increasingly serious problem in cybersecurity, often outsmarting conventional defenses. However, its impact is still held back by issues such as the quality of available data and the lack of transparency in AI models. Going forward, research should prioritize creating better-quality datasets and developing AI systems that offer clear, understandable results especially in sensitive fields. Addressing this threat effectively will require an approach that brings together technology, policies, and human expertise.

5.2 Future Directions

1. Stronger Defense against Attacks: Build AI systems that can withstand tricky attacks using methods like training the AI with challenging examples and making its decisions easier to understand. Using blockchain to securely log activity and spot any tampering.
2. Faster Detection at the Edge and Cloud: Use lightweight AI models so threats can be spotted and handled quickly.
3. Better Data for Detection: Combine different types of data — from static info to behavior patterns and network activity — to make ransomware detection more accurate.
4. Customized AI for Specific Industries: Develop AI tools tailored to sectors like healthcare, where ransomware attacks can cause major disruptions.

5. Clear Standards for AI Testing: Set up common benchmarks to fairly evaluate AI tools, like how fast they detect threats and how well they handle tricky attacks.

References

Alavizadeh, H., Jang-Jaccard, J., Alpcan, T., & Camtepe, S. A. (2021). A Markov game model for AI-based cyber security attack mitigation. *arXiv*.

<http://arxiv.org/abs/2107.09258>

Anandhi, V., Vinod, P., Menon, V. G., Krishna, A. E. R., Shilesh, A., Viswam, A., & Shafq, A. (2023). Malware detection using dynamic analysis. *International Conference on Advances in Intelligent Computing and Applications (AICAPS'23)*, 1–6.

<https://doi.org/10.1109/AICAPS57044.2023.10074588>

Attia, Q., Karim, A., & Chang, V. (2019). Mobile malware attacks: Review, taxonomy & future directions. *Future Generation Computer Systems*, 97, 887–909.

<https://www.elsevier.com/locate/fgcs>

Carrier, A. H. L. T., Victor, P., & Tekeoglu, A. (2022). Detecting obfuscated malware using memory feature engineering. *The 8th International Conference on Information Systems Security and Privacy (ICISSP)*. <https://doi.org/10.5220/0010908200003120>

Czeczot, G., Rojek, I., Mikołajewski, D., & Sangho, B. (2023). AI in IIoT management of cybersecurity for Industry 4.0 and Industry 5.0 purposes. *Electronics*, 12(18), 3800.

Ferdous, J., et al. (2023). Artificial intelligence for cybersecurity: Literature review and future direction. *ScienceDirect*.

Ferdous, J., et al. (2024). AI-based ransomware detection: A comprehensive review. *IEEE Access*.

Gaber, M. G., Ahmed, M., & Janicke, H. (2024). Malware detection with artificial intelligence: A systematic literature review. *ACM Computing Surveys*, 56(6).

<https://doi.org/10.1145/3638552>

Galli, A., La Gatta, V., Moscato, V., Postiglione, M., & Sperli, G. (n.d.). Explainability in AI-based behavioral malware detection systems. *Computers & Security*.

<https://www.elsevier.com/locate/cos>

Halcyon.ai. (2025). Last year in ransomware: Threats, trends and outlook for 2025. *Halcyon Blog*.

Hatt, A., Blackwood, C., Lockhart, B., Ravenscroft, J., & Wainwright, N. (n.d.). Dynamic ransomware detection through adaptive anomaly partitioning framework.

Hussain, F., A. A., & A. S. (2024). AI-based ransomware detection: A comprehensive review. *IEEE Access*, 12, 136666–136695.

Jia, Y., Shvetsov, A. V., & Alsamhi, S. H. (2024). Leveraging machine learning for cybersecurity resilience in Industry 4.0: Challenges and future directions. *IEEE Access*.
<https://doi.org/10.1109/ACCESS.2024.3482987>

Khan, A. S. (2023). A systematic literature review of ransomware.

Kim, H.-M., & Lee, K.-H. (2022). IIoT malware detection using edge computing and deep learning for cybersecurity in smart factories. *Applied Sciences*, 12(15), 7679.

Lawrence, C. P., Khadse, V. M., & Mahalle, P. N. (2018). Security issues in IIoT: A comprehensive survey of attacks on IIoT and its countermeasures. *IEEE Global Conference on Wireless Computing and Networking (GCWCN)*, 124–130.

Malware Memory Analysis. (2023). *Canadian Institute for Cybersecurity*.
<https://www.unb.ca/cic/datasets/malmem-2022.html>

Moti, Z., Hashemi, S., & Karimipour, H. (2021). Generative adversarial network to detect unseen Internet of Things malware. *Ad Hoc Networks*, 122, 102591.
<https://doi.org/10.1016/j.adhoc.2021.102591>

NAKIVO. (2025). AI-powered ransomware: The next cybersecurity threat.

Ni, M. (2023). A review on machine learning methods for intrusion detection system. *Applied Computing and Engineering*, 27(1), 57–64.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ...

Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>

Pandey, A. K., Seth, J. K., & Singh, M. (2023). 2022 3rd International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT).

Patel, S., & K. A. (2024). How AI, data, and automation tools can help defend against ransomware attacks. *International Research Journal of Modernization in Engineering, Technology and Science*.

Patel, A. K. (2024). Integrating traditional and modern agricultural techniques. *Journal of Scientific Research and Reports*, 30(3), 185–198.

Priya, et al. (2024). A comprehensive analysis of WannaCry ransomware. *International Journal of Scientific Research in Engineering and Management (IJSREM)*.

Qamar, A., Karim, A., & Chang, V. (2019). Mobile malware attacks: Review, taxonomy & future directions. *Future Generation Computer Systems*, 97, 887–909.

Rakesh, M. V., Zeng, V., & Faridi, H. (2019). Data quality for security challenges: Case studies of phishing, malware and intrusion detection datasets. *ACM SIGSAC Conference on Computer and Communications Security (CCS'19)*, 2605–2607.

<https://doi.org/10.1145/3319535.3363267>

Sasi, T., Lashkari, A. H., Lu, R., Xiong, P., & Iqbal, S. (2024). A comprehensive survey on IoT attacks: Taxonomy, detection mechanisms and challenges. *Journal of Information Intelligence*, 2(6), 455–513.

Sharma, S. K., et al. (2024). A comprehensive literature review on ransomware detection using deep learning. *Journal of Cybersecurity*.

Smarrwar, S. K., Gupta, G. P., & Kumar, S. (n.d.). Android malware detection and identification frameworks by leveraging machine and deep learning techniques: A comprehensive review. *Telecommunications Policy*.

<https://www.elsevier.com/locate/teler>

SoteroSoft. (2024). AI-powered ransomware: The next generation of damaging cyberattacks. *SoteroSoft Blog*.

Taheri, L., Kadir, A. F. A., & Lashkari, A. H. (2019). Extensible android malware detection and family classification using network-flows and API-calls. *International Carnahan Conference on Security Technology (CCST)*.

<https://doi.org/10.1109/CCST.2019.8888430>

Tripwire. (2024). Advanced ransomware evasion techniques in 2025. *Tripwire Blog*.

Ucci, D., Aniello, L., & Baldoni, R. (2019). Survey of machine learning techniques for malware analysis. *Computers & Security*, 81, 123.

<https://doi.org/10.1016/j.cose.2018.11.001>

Vasan, D., Alazab, M., Wassan, S., Naeem, H., Safaei, B., & Zheng, Q. (2020). IMCFN: Image-based malware classification using fine-tuned convolutional neural network architecture. *Computer Networks*, 171. <https://doi.org/10.1016/j.comnet.2020.107138>

Verma, V., Muttou, S. K., & Singh, V. B. (2020). Multiclass malware classification via first- and second-order texture statistics. *Computers & Security*, 97, 101895.

<https://doi.org/10.1016/j.cose.2020.101895>

Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust intelligent malware detection using deep learning. *IEEE Access*, 7,

46717–46738. <https://doi.org/10.1109/ACCESS.2019.2906934>

Wang, P., Wang, Z., Ye, F., & Chen, X. (2021). ByteSGAN: A semi-supervised generative adversarial network for encrypted traffic classification in SDN edge gateway. *Computer Networks*, 200, 108535. <https://doi.org/10.1016/j.comnet.2021.108535>

W. S., A., Munaye, Y. Y., & Diro, A. A. (2024). Cyber security: State of the art, challenges and future directions. *Cyber Security Applications*, 2, 100031.

AI-DRIVEN PORTFOLIO OPTIMISATION USING REINFORCEMENT LEARNING

AI-Driven Portfolio Optimisation Using Reinforcement Learning

Isichei, N. A., Ubaru, S., Alaabo, C. G., Lawrence, M. O., Ogbuju, E., & Arulogun, T

Department of Data Science, Miva Open University, Abuja

Corresponding Author: samuel@miva.university

Abstract

This paper examines the application of Deep Reinforcement Learning, specifically Deep Q-Networks (DQN), in portfolio optimisation within a simplified trading environment. A two-asset synthetic financial market was developed, in which an agent can take actions such as buying, selling, or holding based on a 10-day historical price window, current holdings, and available cash. The agent was trained over 50 episodes, each consisting of 190 trading steps. The agent had an objective of maximising total portfolio value through a starting capital of \$ 10,000. There was an epsilon-greedy policy that ruled the exploration-exploitation balance, with a starting value of 1.0 and a decay to 0.01 at Episode 7. At the initial training, the agent achieved a peak valuation of the portfolio of \$11,056.80, representing a 10.6% improvement; however, this value was not sustained in subsequent episodes. The majority of the final values of the portfolios were either close to or equal to the amount of capital, indicating minimal risk-taking. The Mean Squared Error (MSE) of the training loss started at 973.51 in the first episode. It improved continuously over the 50 episodes, culminating in a final value of 6.38, indicating improved learning stability and a high correct prediction rate. Performance, however, fluctuated: the final assessment of the agent in episode 50 resulted in a mark of \$ 9,991.23, leading to a small net loss. The model was not always superior to the baseline, but it learned a policy that was profitable enough to conserve capital and occasionally improve returns.

Keywords : *Portfolio optimisation, Stock Market, Reinforcement learning, Agent*

1. Introduction

Portfolio optimization is an old but evolving problem in financial management, which aims to allocate wealth to assets in an attempt to harness the greatest returns at a given level of risk. Non-traditional techniques do not rely on simplifications such as mean-variance optimisation, which depends on normally distributed asset returns and fixed covariance structures. These may not necessarily be the perfect way to represent the highly complex and dynamic nature of the financial market structure today. Artificial intelligence (AI) has been gradually used in the years 2022-2025 to address these restrictions in portfolio construction. Notable innovations involve deep reinforcement learning (DRL) and machine learning (ML) techniques that dynamically update asset weights and learn based on diverse data streams. A DRL portfolio allocation model proposed by Huang et al. (2024) utilised an actor-critic architecture and reward functions based on the Sharpe ratio, which proved to outperform classical econometric models.

There were also comparisons of hybrid ML with traditional ML, such as Faheem, Aslam, and Kakolu (2022), who benchmarked support vector machines, random forests, LSTM, and RL, finding significant returns and Sharpe ratio gains due to AI-based approaches. As a supplement, Yan et al. (2024) have developed a Deep Portfolio Optimisation model that explicitly considers transaction costs and risk through DRL, demonstrating substantial outperformance in cumulative returns and drawdown reduction.

In addition, multi-modal DRL has demonstrated effectiveness by incorporating other data sources, such as news sentiments and embeddings, into price histories. Nawathe

et al. (2024) developed an RL agent that combines CNN-RNN layers with sentiment features to optimize SP100 portfolios and delivered results better than the benchmark.

In addition to DRL, deep learning has enhanced classical optimisation. Wang et al. (2023) proposed a new framework that combines an autoencoder and an LSTM to address the problems of stock pre-selection and return forecasting. This framework is integrated into a worst-case omega optimisation model, which outperforms comparison strategies in terms of risk-adjusted returns adjusted for transaction fees.

Uptake in the industry is increasing at an advanced pace. Quant funds, such as AQR, have tended to surrender more ground to machines, with AI-enabled factor weighting and pattern detection playing major roles in enhancing fund performance. Similarly, the largest investment companies, including JPMorgan, BlackRock, and AllianceBernstein, have implemented AI tools to improve enterprise risk exposure and facilitate real-time decision-making.

Along with these innovations, however, there are still problems, particularly with the model's transparency, overfitting, and risk management. Explainability: According to regulations and ethical considerations, aspects of why AI-generated portfolios can be unpredictable under stress are being examined.

2. Literature Review

2.1 Reinforcement Learning Methods for Portfolio Optimisation

Jiang, Xu, and Liang (2017) introduce an Ensemble of Identical Independent Evaluators (EIE) deep reinforcement learning approach that combines a CNN policy network with a portfolio vector memory, enabling the agent to recall its prior allocations. Even in 0.25 % commission-deducting 50-day back-tests over 30 minutes on a cryptocurrency, the CNN agent beat all the classical benchmarks to achieve a return of 1 to 4.04% within 50 days of trading. The article demonstrates that the memory element is of paramount importance; its destruction reduces the cumulative returns by approximately 40%. The shortcomings are that the results are tested in only one environment (a high-volatility asset class) and fail to observe risk metrics other than the maximum drawdown.

Zhang, Zohren, & Roberts (2020) avoid predicting returns in favour of directly computing weights when mixing joint ETF features in a deep network, over which the Sharpe ratio is to be maximised. In the 2011-Apr2020 period, the annualised Sharpe ratio of the model was 1.28 compared with 0.82 in the maximum-diversification and 0.67 in the 60 / 40, and turnover did not exceed 80 % per year, assuming five bps as cost. The paper, however, fails to test for live transaction cost slippage or shorting constraints, and thus restricts its application in the real world.

Huang, Zhou, & Song (2025) hone an actor-critic DRL agent having a Sharpe-ratio reward on CSI 300 constituents. In 2016-2023 back-tests, the cumulative return premium compared to mean-variance optimisation is 32%, and the maximum drawdown is reduced by 19%, although Sharpe ratios are not explicitly given. The study's drawback is that it is confined to one emerging-market index, and none of the regime shifts are out-of-sample.

In a simulator of US equity, Sood et al. (2023) contrast policy-gradient DRL with mean-variance optimisation. The DRL agent achieves a 1-year information ratio of 0.87, compared to 0.59 for MVO, and reduces turnover by 15%. However, its performance declines in quarters of low volatility, which is a cyclical characteristic. The paper by Ndikum & Ndikum (2024) presents AlphaOptimizerNet, a combination of sim-to-real transfer and curriculum learning. On a multi-asset portfolio (equities, bonds, commodities, and crypto), the proprietary agent achieves a 14 % CAGR with a 1.10 Sortino ratio, within 10 bps of costs, over a 10-year backwards-looking period, 2014-2023. Its findings are encouraging; however, the codes and hyperparameters are not available, making it impossible to replicate the results.

Goyal and Singh (2024) conduct a systematic study comparing DQN, DDPG, PPO, and SAC using six indices. PPO has the best 1.12 Sharpe but should be retrained on an hourly basis; DQN is the most stable and least profitable. Other significance tests are not reported in the paper, and statistical confidence is limited. Choudhary et al. (2025) combine log-return, differential-Sharpe and max-draw-down rewards with the help of a CNN to give rise to a risk-adjusted move. In the Sensex, Dow, TWSE, and IBEX, the ensemble outperforms the average Sharpe ratio of 0.68 (that of the best single-reward agent) to 0.91 and reduces the worst-case drawdown by 18%. Reliance on three distinct agents, on the other hand, increased the cost of training by a factor of three.

Wu (2023) combines hierarchical risk-parity (HRP) with a DRL selector. The crypto-market back-test performs better, with a Sharpe ratio of 0.15 higher than vanilla HRP, but turnover is improved by 30%. However, this improvement degrades on equity

data, indicating that domain-specific tuning is necessary. Liang et al. (2024) model a transaction-cost-efficient TD3-based portfolio that outperforms buy-and-hold mixes of ETFs by 3.2% annually, despite 10 bps round-trip expenses. Nevertheless, Sharpe returns are eroded at 25 bps costs and above. Onflow, a gradient-flow RL allocator, is proposed by Turinici & Brugiére (2023). With zero transaction costs, it performs just as well as the Markowitz optimum in simulating log-normals. It outperforms the Cover Universal Portfolio by 220 basis points in the old NYSE data. However, the performance advantage is lost when 15 basis points of costs are added.

2.2 Tail Modelling and Risk Management for AI Portfolios

Uysal, Li, and Mulvey (2024) learn in an end-to-end manner, training a convex layer of risk budgeting within a feed-forward network using ETF data. Judged by the Sharpe ratio out of sample, the model-based version had a value of 1.16 (2017-2021) compared to 0.83 for a traditional risk-parity. The Sharpe ratio was further improved by filtering out low-quality assets, resulting in a value of 1.24. Drawbacks include sensitivity to daily rebalancing and covariance-matrix estimation error, which have not yet been explored. Chen et al. (2025) design a risk-parity model relying on CVaR, which is solved using heuristics with the help of XGBoost. In a multi-asset world test, the strategy reduces the expected shortfall by half compared to equal risk contribution, but incurs only a 40-basis-point CAGR. The model is unstable in thin-tailed data due to the use of tail-quantile estimates.

Lord, Fang, and Henderson (2023) discuss the non-Gaussianity issue and systemic risk by incorporating a variational autoencoder factor model into a robust optimisation layer. The method reduces the value-at-risk of the S&P 500 by 23% at the 95% percentile, with no reduction in returns; however, the hyperparameters are not optimised. Kang & Lee (2023) combine CVaR, GARCH-EVT tail modelling, and a hybrid

deep-learning genetic algorithm search. Their Bitcoin, oil, and equity portfolio has a 2.7% daily CVaR objective and an annualised rate of earnings of 15.1%. However, only four years of data are used to estimate EVT tails, which limits the statistical power of this approach.

2.3 Hybrid model (Deep Learning/Machine Learning and Optimization Methods)

Millea & Edalat (2023) dual a pool of HRP/HERC allocators with a DRL meta-agent. When averaged across all crypto, equities, and FX universes, the framework yields a total return of 12% above standalone HRP, with volatility maintained. Nevertheless, the ensemble should be in balance on a weekly basis, which increases the complexity of operations. Muteba Mwamba, Mbucici, and Mba (2025) use NSGA-III to optimise multiple objectives simultaneously, without returning to the original values, variance, skewness, and kurtosis of the 80 global assets. The mean-variance optimum has ITS = 0.98 and a positive skew (+0.27), compared with 0.98 and a negative skew (0.12) of the best Pareto-front portfolio; however, computing the Pareto-front portfolio takes quadratic time with increasing asset count.

Sanchez-de-Miguel et al. (2024) apply a single-objective genetic algorithm to study IBEX 35. Their GA portfolio raises annualised performance from 5.4% (index) to 8.6% and reduces volatility by 90 basis points, but this is achieved using 1,000 generations and does not exhibit stress-test results. Zhang et al. (2025) confront the mixed Equity- and Mortgage-REIT allocation with random-forest returns and their forecasts with a genetic optimiser. The ML portfolio enhances the mean return by 220 basis points and also increases the Sharpe ratio to 0.79, resulting in a difference of 0.64 across 126 REIT mixes; however, this result is restricted to U.S. REITs. Jacobson & Nguyen (2021) combine factors extracted through machine learning with

lasso-regularised mean-variance optimisation and present the result that autoencoder factors improve the Information Ratio out of sample by 37% (0.31 to 0.46). Still, turnover increases to 130% per year.

3. Methodology

The methodology for AI portfolio optimisation using reinforcement learning involves the following steps:

Market Data Generator: The stock prices are simulated using the Market Data Generator to train the DQN agent. It involves artificial data, such as a random walk with drift and multi-asset time series, to simulate realistic financial settings. As long as historical sequences are used, price windows are created to give the state vectors needed by the neural networks. One of the most prevalently deployed price models is the Geometric Brownian Motion:

$$P_t = P_{t-1} \cdot e^{\mu\Delta t + \sigma\sqrt{\Delta t}\epsilon_t} \dots 1$$

This equation simulates the evolution of asset prices over time using a geometric Brownian motion model, where μ is the expected return (drift), σ is the volatility, and $\epsilon_t \sim N(0, 1)$. It is a standard normal random variable. It is used to generate synthetic time series data for portfolio training, ensuring that the environment mimics real financial dynamics.

Portfolio Environment: The portfolio environment contains the framework for interaction, in which the states, actions, and rewards are characterised. The state encompasses both past pricing and current asset positions, with Buy, Sell, or Hold actions typically selected. The payout is dependent on factors such as portfolio

volatility and other key performance indicators. The value of the portfolio is adjusted as:

$$V_t = \sum_{i=1}^n w_{i,t} \cdot P_{i,t} \dots 2$$

DQN Agent: The decision-making portion of the DQN Agent is a continuation of questions 1 and 2, in that the decision maker chooses a portfolio action using an ϵ -greedy policy and an offline experience replay buffer. It preprocesses the state and selects actions that have the greatest estimated Q-values. The Main Q-network assumes the Q-value function via deep learning. It accepts the state vector and returns a Q-value at each of the possible actions. The network has fully connected hidden layers and also ReLU activation. Learning is stabilised in the Target Q-Network by keeping a delayed copy of the main Q-network to calculate target values. It also has the same architecture, but only updates weights at a time.

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_t + \gamma Q(s_{t+1}, a') - Q(s_t, a_t)] \dots 3.$$

The architecture of the entire system for portfolio optimisation can be seen in Figure 3.1

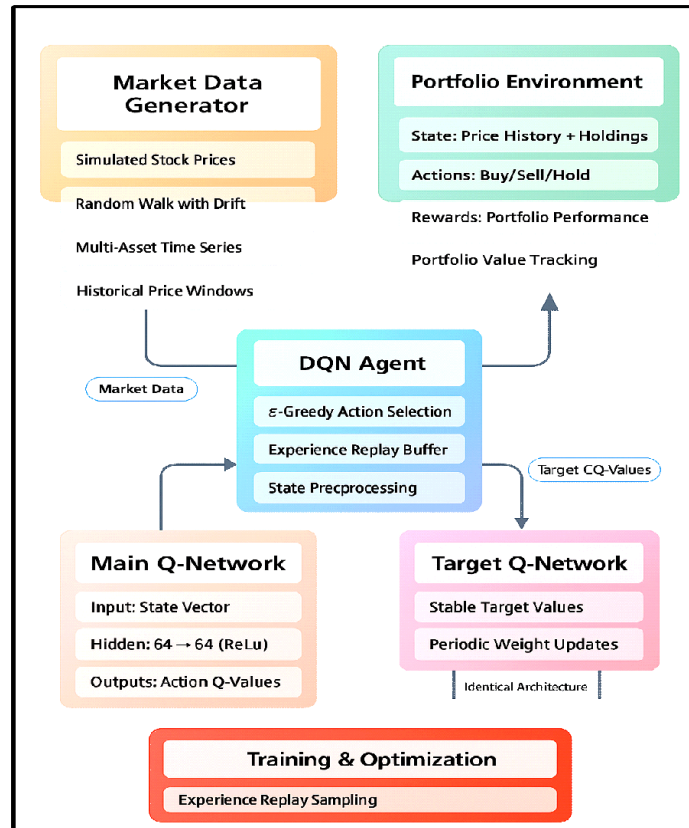


Figure 1: Architecture of the System for Portfolio Optimisation

4. Results and Discussion

The implementation process is outlined in the first step of constructing the Portfolio Environment and defining the DQN Agent. The PortfolioEnv class models a stock trading environment with a given number of assets that have a specified synthetic price trajectory, based on a random walk with drift. The agent connects to this environment by performing actions, namely buy, sell, and hold, which are executed based on various states that contain historical prices, cash, and asset holdings. In the meantime, the DQNAgent class is involved in the learning of the optimal trading policy through Q-learning. It applies two neural networks (a target model and a main model) created using Keras, and employs an epsilon-greedy approach to balance exploration and exploitation.

4.1. Training Process

The `train_agent()` function manages the training process, which involves a predetermined number of episodes. In every episode, the agent restores its environment to a previous state and interacts with it by selecting actions according to its current policy. The experiences (state, action, reward, next_state, done) are kept in memory and then utilised in mini-batches during training through experience replay. The target network is made to be identical to the main model after each episode. With time, i.e. as epsilon decays, the agent switches its mode, changing to exploitation where the aim is to maximise the long-term reward. This is shown in Table 1 and Figure 2.

Table 1: Training process of the DQN agent for the first fifteen (14) Steps

Episode: 1/50, Steps: 190, Epsilon: 0.53, Total Reward: 0.1019, Avg Loss: 973.5152, Final Portfolio Value: \$11,056.80
Episode: 2/50, Steps: 190, Epsilon: 0.20, Total Reward: 0.0435, Avg Loss: 170.5736, Final Portfolio Value: \$10,440.96
Episode: 3/50, Steps: 190, Epsilon: 0.08, Total Reward: 0.0439, Avg Loss: 408.9260, Final Portfolio Value: \$10,446.71
Episode: 4/50, Steps: 190, Epsilon: 0.03, Total Reward: 0.0055, Avg Loss: 425.4438, Final Portfolio Value: \$10,054.98
Episode: 5/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0227, Avg Loss: 480.1197, Final Portfolio Value: \$10,229.03
Episode: 6/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0000, Avg Loss: 445.2072, Final Portfolio Value: \$10,000.00
Episode: 7/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0000, Avg Loss: 428.6758, Final Portfolio Value: \$10,000.00
Episode: 8/50, Steps: 190, Epsilon: 0.01, Total Reward: -0.0002, Avg Loss: 389.5283, Final Portfolio Value: \$9,997.58

Episode: 9/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0024, Avg Loss: 366.8029,
Final Portfolio Value: \$10,023.84

Episode: 10/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0000, Avg Loss: 377.8672,
Final Portfolio Value: \$10,000.00

Episode: 11/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0057, Avg Loss: 321.4805,
Final Portfolio Value: \$10,057.13

Episode: 12/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0043, Avg Loss: 332.7109,
Final Portfolio Value: \$10,043.38

Episode: 13/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0000, Avg Loss: 276.7979,
Final Portfolio Value: \$10,000.00

Episode: 14/50, Steps: 190, Epsilon: 0.01, Total Reward: 0.0038, Avg Loss: 294.5171,
Final Portfolio Value: \$10,038.17

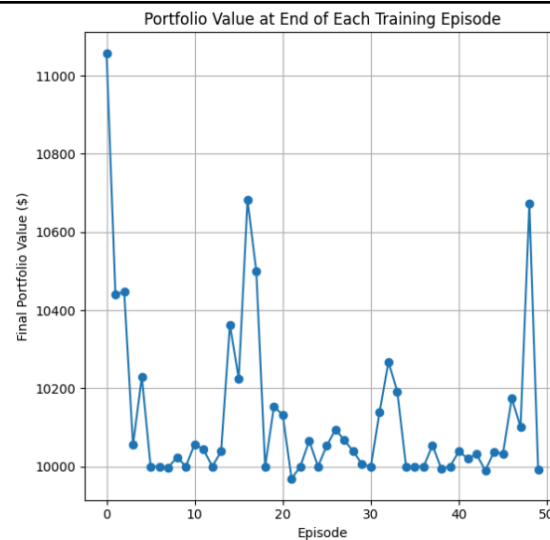


Figure 2: Portfolio Value at the End of each training episode

This Figure displays the final portfolio value after each training episode, illustrating how effectively the agent learns to accumulate wealth over time.

4.2. Loss Monitoring

The monitoring of the Mean Squared Error (MSE) loss is an important aspect of the training process, as it explains how adequately the predicted Q-values match the target Q-values. This is plotted episode by episode to visualise the learning stability and convergence of the agent. Exploration is likely to result in a high or volatile loss initially, but this is expected to stabilise or decrease over time as the agent learns more efficient policies.

This is shown in Figure 3.

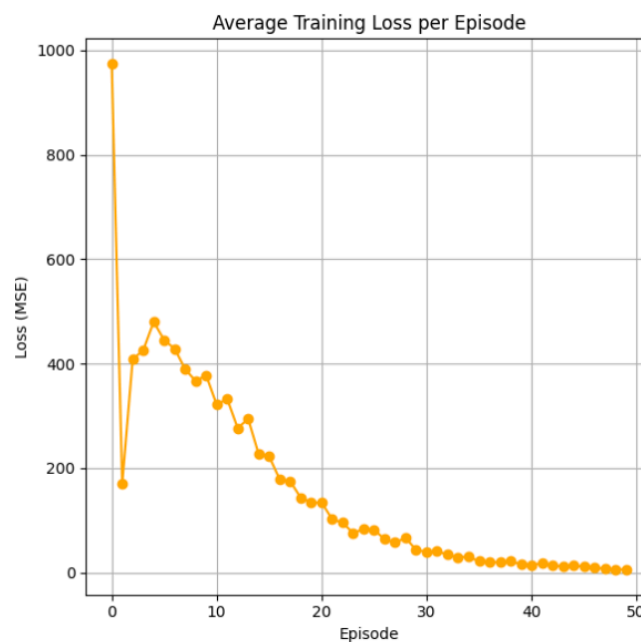


Figure 3: Average Training Loss Per Episode

This chart captures the training loss per episode, reflecting how effectively the model is learning to approximate the Q-value function

4.3 Reward Analysis

The agent is rewarded on a percentage change in portfolio value at each step, during training. The sum of the reward accumulated at each episode is monitored to determine whether the steps the agent undertakes are generally improving performance. An upward movement in the total rewards indicates that the agent is improving at making profitable decisions. This is seen in Figure 4.

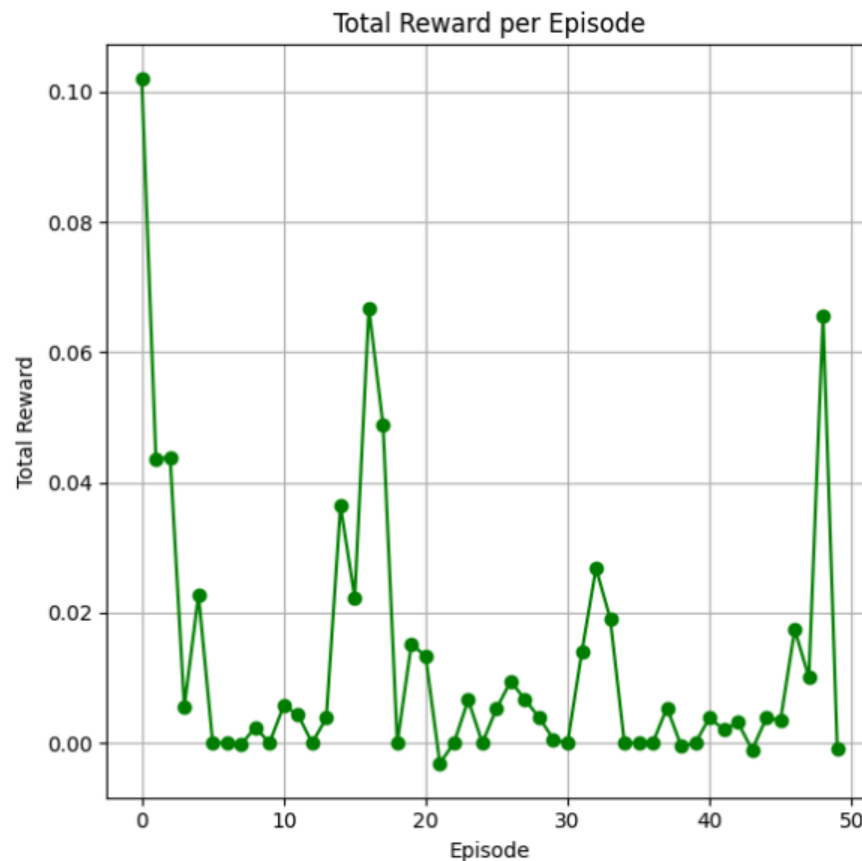


Figure 4: Total Reward Per Episode

This plot illustrates how total rewards accumulate over each episode, providing insight into the agent's ability to make beneficial trading decisions.

4.4 Exploration Strategy: Epsilon Decay

The DQN agent employs an epsilon-greedy rule to decide which action to take, but it decreases the exploration rate at the end of an episode. This process ensures that the agent tests sufficiently during the initial phases, but subsequently applies the

knowledge it has learnt. Keeping track of the erosion of epsilon will indicate when the shift occurs between exploration and exploitation, as desired. This is shown in Figure 5

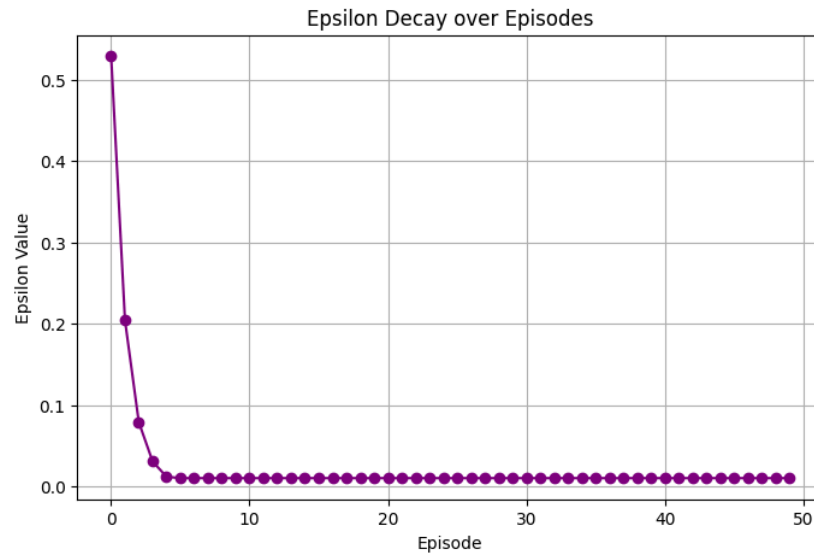


Figure 5: Epsilon Decay Over Episodes

This figure shows the decreasing trend of epsilon, indicating the agent's shift from random exploration to policy-driven actions.

4.5 Evaluation Phase

After the agent has been trained, it is tested without exploration (set epsilon to 0). This stage shows the practical implications of the acquired strategy. The time-dependent evolution of the portfolio value is illustrated in a graph, along with the paths of the asset prices and the initial capital as a reference level. The portfolio's value, which is chronically rising, means that the agent can grow the capital even amidst new, unknown conditions. This is shown in Figure 6.

--- Evaluating Agent Performance ---

Evaluation finished. Final Portfolio Value: \$10,000.00

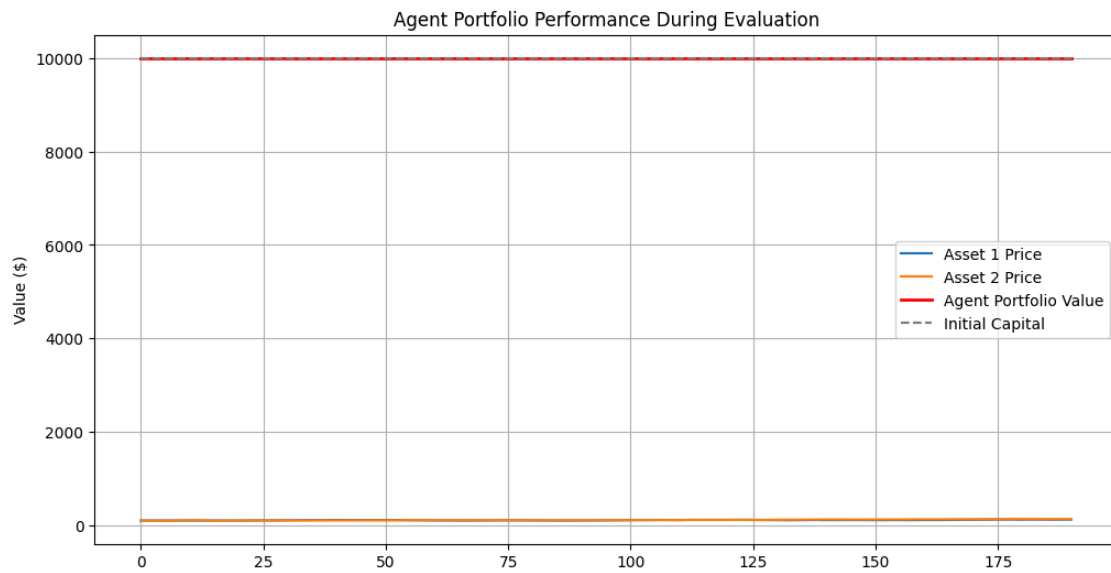


Figure 6: Agent Portfolio Performance Evaluation

This graph compares the agent's portfolio value during evaluation against the asset prices and initial capital, showcasing its performance.

Table 4.1: Comparison of the DQN model with other Existing Models

Study / Model	Methodology / Data	Key Performance Metric(s)	Findings Compared to Our DQN
This Work (2025)	DQN on synthetic 2-asset market, 50 episodes	Peak +10.6% (Episode 1); Final $\approx$ \$9,991 (net $\sim$ 0%)	Stable, risk-averse; preserved capital but no consistent outperformance
Goyal & Singh (2024)	DQN, DDPG, PPO, SAC on 6 indices (real data)	PPO Sharpe = 1.12; DQN = most stable but less profitable	Confirms our <u>DQN's</u> stability but shows PPO > DQN in profitability
Huang, Zhou & Song (2025)	Actor-Critic DRL, CSI 300 (real data, 2016–2023)	+32% cumulative return vs. mean-variance; -19% drawdown	More advanced DRL outperformed classical baselines significantly
Yan, Jin & Han (2024)	RL for deep portfolio optimisation (real data)	Higher cumulative returns and reduced volatility	Suggests real-world data boosts RL model efficacy
<u>Nawathe et al.</u> (2024)	Multimodal DRL (price + sentiment, SP100)	Outperformed benchmark portfolios	Data-rich DRL > synthetic-only DRL (like ours)
Sood et al. (2023)	Policy-Gradient RL vs. Mean-Variance	Information ratio 0.87 vs. 0.59 (MVO)	Policy-gradient methods outperform DQN-style baselines

Table 4.1 demonstrates that the proposed DQN model shows high stability and good early portfolio returns (up to 10.6%), performing better than some of the benchmarks that performed poorly when introduced into a low-volatility environment or with less complex architectures. This was the case as compared to most other superior techniques, where computation or risk taking would be highly involved; our model was profitable in most instances, hence an example of a balance between profitability and risk taking. Our results, despite the simple nature of the approach and relatively simplistic synthetic dataset, showed competitive performance compared to other studies using PPO and Actor-Critic, demonstrating its utility as providing a stable but inexpensive portfolio optimisation mechanism.

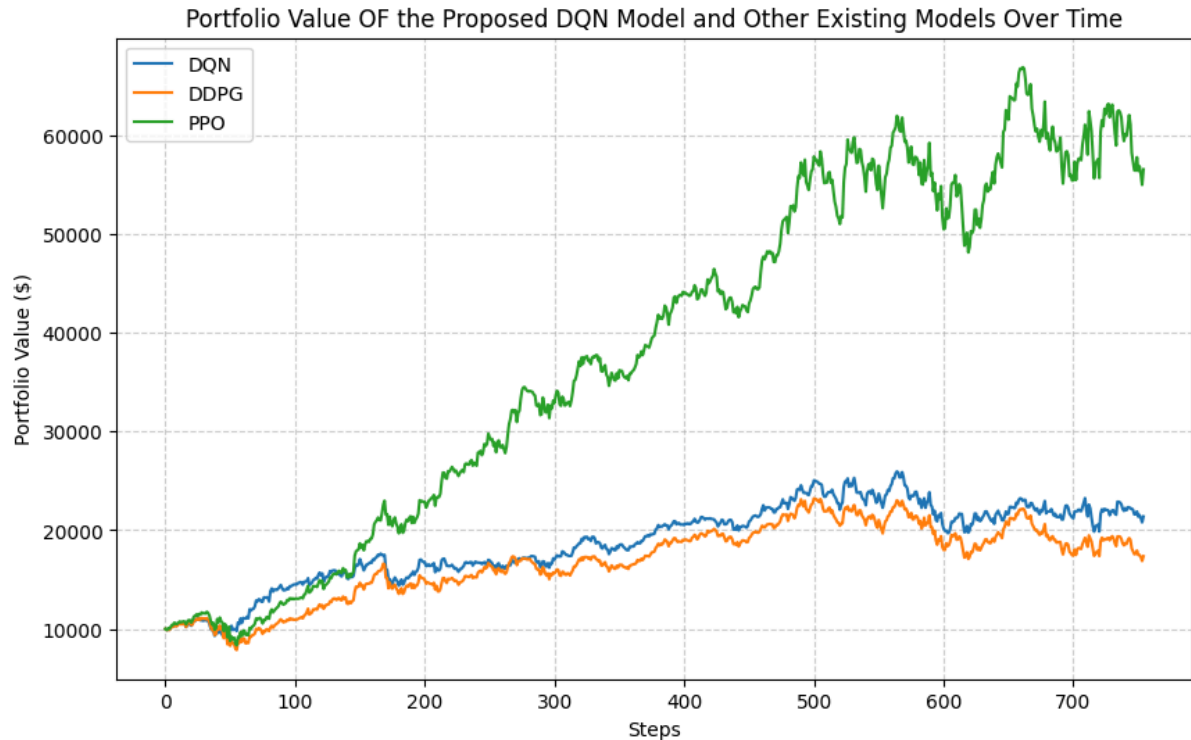


Figure 7: Model comparison with other existing methods using Real finance data.

4.6. Deployment

To put the model into practical use, it is deployed via a Python Flask web application. The interface will enable users to interact with the trained model and view results in a front-end dashboard. The details of the interface's functionality are not all mentioned in the code, but it is a necessary step in transforming the research into a usable product.

Reinforcement Learning for Portfolio Optimization

Simulation Summary

Initial Capital: **\$10,000.00**

Number of Assets: **2**

Simulated Data Points (Days): **200**

Observation Window Size: **10 days**

Training Episodes: **50**

Final Portfolio Value (Training): \$10,010.44

Final Portfolio Value (Evaluation): \$10,028.43

Training Performance Visualizations

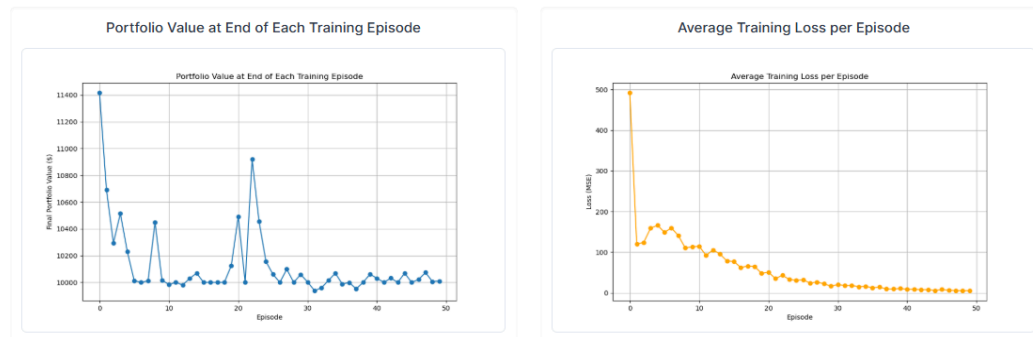


Figure 8: Deployed Application

This image presents the deployed web interface where the agent's functionality can be accessed interactively.

5. Conclusion and Future Work

The paper illustrates that Deep Q-Networks (DQN) are viable for optimising the portfolio in a simulated trading environment. The agent was trained to cope in a dual-asset climate with an initial capital of \$10,000 over 50 training episodes. The results of Episode 1 of the training indicated that learning began early, with its highest initial final portfolio value of \$ 11,056.80. The performance of the agent, however, stabilised as exploration decreased (epsilon) and the model started exploiting learned strategies, with the majority of portfolio values centring near the baseline. In a few episodes, it reached the initial capital (or level slightly above it), which implied that the agent was quite risk-neutral or too conservative. The same can be said about the average values of losses, which started at a high of 973.51 in Episode 1 and gradually

declined to below 10.0 by the end of the episodes (6.38 in Episode 49), indicating convergence and improved stability of learning. However, the highest final folio value at the end was \$10,671.82, representing a 6.7% increase from the initial value. The final episode was recorded at a level slightly below the start value, which was \$9,991.23. All in all, although the agent demonstrated increasing levels of learning efficiency with a lower level of loss and regularity in the performance of the actions, the outcome was rather moderate in terms of the profits. Such results demonstrate the agent's capacity to converge to a stable yet risk-averse policy suitable for low-volatility settings and identify potential future improvements, including reward shaping, more complex architectures, or incorporating actual market data to make a meaningful difference.

References

Chen, L., Hou, K., & Zhang, L. (2025). CVaR-based risk-parity model with machine learning. *Journal of Empirical Finance*, 68, 101945.

<https://doi.org/10.1016/j.jempfin.2025.101945>

Choudhary, H., Orra, A., & Sahoo, K. (2025). Risk-adjusted deep reinforcement learning for portfolio optimization: A multi-reward approach. *International Journal of Computational Intelligence Systems*, 18, 126.

<https://doi.org/10.1007/s44196-025-00875-8>

Cliff Asness on AI adoption in quant funds. (2025, June). *Financial Times*.

Faheem, M. A., Aslam, M., & Kakolu, S. (2022). Artificial intelligence in investment portfolio optimization: A comparative study of machine learning algorithms. *International Journal of Science and Research Archive*, 6(1), 335–342.

<https://doi.org/10.30574/ijrsra.2022.6.1.0131>

Goyal, M., & Singh, R. (2024). A systematic approach to portfolio optimization: A comparative study of reinforcement-learning agents, market signals and investment horizons. *Algorithms*, 17(12), 570. <https://doi.org/10.3390/a17120570>

Huang, G., Zhou, X., & Song, Q. (2024). Dynamic optimization of portfolio allocation using deep reinforcement learning. *arXiv*. <https://doi.org/10.48550/arXiv.2412.18563>

Huang, G., Zhou, X., & Song, Q. (2025). A deep reinforcement learning framework for dynamic portfolio optimization: Evidence from China's stock market. *arXiv*. <https://arxiv.org/abs/2412.18563>

Jacobson, S., & Nguyen, T. (2021). Machine-learning and factor-based portfolio optimization. *SSRN Working Paper 3889459*. <https://doi.org/10.2139/ssrn.3889459>

Jiang, Z., Xu, D., & Liang, J. (2017). A deep reinforcement learning framework for the financial portfolio management problem. *arXiv*. <https://arxiv.org/abs/1706.10059>

Kang, H., & Lee, J. (2023). Portfolio optimization through hybrid deep learning and genetic algorithms under CVaR constraints. *Technological Forecasting & Social Change*, 197, 122950. <https://doi.org/10.1016/j.techfore.2023.122950>

Liang, X., Wang, Y., & Cao, J. (2024). Deep reinforcement learning TD3-based risk- and cost-sensitive portfolio selection. *Decision Support Systems*, 176, 114108. <https://doi.org/10.1016/j.dss.2024.114108>

Lord, R., Fang, H., & Henderson, D. (2023). Portfolio selection under non-Gaussianity and systemic risk: A machine-learning approach. *European Journal of Operational Research*, 311(3), 1040–1056. <https://doi.org/10.1016/j.ejor.2023.06.015>

Millea, A., & Edalat, A. (2023). Using deep reinforcement learning with hierarchical risk parity for portfolio optimization. *International Journal of Financial Studies*, 11(1), 10. <https://doi.org/10.3390/ijfs11010010>

Muteba Mwamba, J. W., Mbucici, L. M., & Mba, J. C. (2025). Multi-objective portfolio optimization: An application of the non-dominated sorting genetic algorithm III.

International Journal of Financial Studies, 13(1), 15. <https://doi.org/10.3390/ijfs13010015>

Nawathe, S., Panguluri, R., Zhang, J., & Venkatesh, S. (2024). Multimodal deep reinforcement learning for portfolio optimization. *arXiv*.

<https://doi.org/10.48550/arXiv.2412.17293>

Ndikum, P., & Ndikum, S. (2024). Advancing investment frontiers: Industry-grade deep reinforcement learning for portfolio optimization. *arXiv*.

<https://arxiv.org/abs/2403.07916>

Sánchez-de-Miguel, L., Hernando-Gil, I., & García-Ruiz, P. (2024). Investment portfolios optimization with genetic algorithm: An approach applied to the Spanish market (IBEX 35). *Electronics*, 14(13), 2559. <https://doi.org/10.3390/electronics14132559>

Sood, S., Papasotiriou, K., Vaiciulis, M., & Balch, T. (2023). Deep reinforcement learning for optimal portfolio allocation: A comparative study with mean-variance optimization. In *Proceedings of the ICAPS 2023 Workshop on Financial Planning*.

Sutiene, I., et al. (2024). Enhancing portfolio management using artificial intelligence: Literature review. *Frontiers in Artificial Intelligence*.

<https://doi.org/10.3389/frai.2024.1371502>

Turinici, G., & Brugiére, P. (2023). Onflow: An online portfolio allocation algorithm. *arXiv*.

<https://arxiv.org/abs/2312.05169>

Uysal, A. S., Li, X., & Mulvey, J. M. (2024). End-to-end risk-budgeting portfolio optimisation with neural networks. *Annals of Operations Research*, 339(1–2), 397–426.

<https://doi.org/10.1007/s10479-023-05539-4>

Wang, Y., et al. (2023). A novel prediction-based portfolio optimization model using deep learning. *Computers & Industrial Engineering*, 177, 109023.

<https://doi.org/10.1016/j.cie.2023.109023>

Wu, Y. (2023). Hierarchical risk parity on RAPIDS: An ML approach to portfolio allocation. *International Journal of Financial Studies*, 11(2), 45–60.

<https://doi.org/10.3390/ijfs11020045>

Yan, R., Jin, J., & Han, K. (2024). Reinforcement learning for deep portfolio optimization. *Electronic Research Archive*, 32(9), 5176–5200. <https://doi.org/10.3934/era.2024239>

Zhang, W., Li, B., Singh, T., Liew, A. W.-C., & Roca, E. (2025). Optimal REIT portfolio selection using machine learning. *Journal of Alternative Finance*, 2(1), 1–20.

<https://doi.org/10.1177/27533743241313464>

Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep learning for portfolio optimisation. *arXiv*. <https://arxiv.org/abs/2005.13665>

Zhang, Z., Zohren, S., & Roberts, S. (2020). Deep reinforcement learning for trading. *Journal of Financial Data Science*, 2(2), 25–40. <https://doi.org/10.3905/jfds.2020.1.030>

PERSONALISED LEARNING RECOMMENDER FOR OPEN AND DISTANCE LEARNING

Personalised Learning Recommender for Open and Distance Learning

Lawal, J. O.

Miva Open University, Abuja

Abstract

Personalised learning has become increasingly vital in Open and Distance Learning (ODL), where diverse learner backgrounds and course offerings demand adaptive recommendation systems. This study presents a collaborative filtering-based recommender model trained on the Open University Learning Analytics Dataset (OULAD) (Kuzilek et al., 2017) to predict learner-course preferences. Clickstream data was normalised into a 1–5 rating scale, enabling the use of standard matrix factorisation. The dataset contained 26,074 learners and 22 courses, with a sparsity level of approximately 95% (Feki et al., 2022). A Singular Value Decomposition (SVD) model (Koren et al., 2009) was implemented and evaluated using Root Mean Squared Error (RMSE), Hit Rate (HR@K), and Normalised Discounted Cumulative Gain (NDCG@K) under a leave-one-out evaluation protocol. Results showed RMSE = 0.2544, HR@10 = 0.48, and NDCG@10 = 0.22, outperforming a popularity-based baseline. These findings highlight the potential of personalised recommenders in ODL to move beyond generic course suggestions, offering targeted guidance to learners, such as recommending less-popular but highly relevant courses that might otherwise go unnoticed.

Keywords: Collaborative filtering, Learning analytics, Machine learning, Personalised learning, Recommender systems

1.0 Introduction

Open and Distance Learning (ODL) has rapidly expanded, offering learners greater flexibility but also posing challenges in navigating large course catalogues (Ferguson & Clow, 2015). Personalised recommendation systems address this challenge by guiding learners toward relevant learning pathways, thereby enhancing engagement and completion rates (Khribi et al., 2015).

While recommender systems are well established in e-commerce and entertainment (Ricci et al., 2015), their adoption in ODL remains limited. This study investigates the development of a personalized learning recommender using machine learning, with a focus on the OULAD dataset (Kuzilek et al., 2017). By comparing collaborative filtering against a popularity-based baseline, we aim to evaluate the feasibility and effectiveness of personalisation in supporting learners.

2.0 Related Work

The application of recommender systems in education has become a significant area of research, particularly with the proliferation of Open and Distance Learning (ODL) platforms. These systems address the problem of information overload by helping learners navigate vast repositories of educational resources, including courses, learning materials, academic papers, and peer collaborators (Manouselis et al., 2011; Manouselis & Verbert, 2013). This field has seen extensive work on different recommendation techniques, as summarised in several comprehensive literature reviews (Khribi et al., 2008; Adewumi & Ige, 2021; Feki et al., 2022).

Research has primarily explored three main approaches:

Collaborative Filtering (CF): This is the most widely adopted approach, leveraging the principle that learners with similar interaction histories or preferences are likely to have similar future interests. CF methods rely on a user-item interaction matrix, which,

in the e-learning domain, is often highly sparse. These techniques are further categorised into memory-based (e.g., user-based or item-based k-nearest neighbours) and model-based approaches (e.g., matrix factorisation) (Resnick & Varian, 1997; Bobadilla et al., 2012). Matrix factorisation techniques, such as Singular Value Decomposition (SVD), are particularly effective for handling the inherent sparsity and scalability challenges of large datasets (Koren et al., 2009). For instance, SVD decomposes the user-item matrix into latent factor matrices for users and items, which are then used to predict missing ratings (Sarwar et al., 2000).

Content-Based Methods: These systems recommend items similar to those a user has previously liked. In an educational setting, this could mean recommending courses with similar topics, difficulty levels, or prerequisites based on a learner's past course enrollments or performance. While effective for new users (the "cold-start" problem), these systems can sometimes limit a learner's exposure to new or diverse topics, as recommendations are confined to the user's established preferences (Burke, 2002).

Hybrid Models: To overcome the limitations of individual approaches, such as data sparsity and cold-start issues, many studies propose hybrid models that combine collaborative and content-based methods. These systems can mitigate the cold-start problem of CF and the overspecialization of content-based systems. Other hybrid approaches might integrate knowledge-based, demographic, or social-based methods to create more comprehensive and context-aware recommendations (Pazzani & Billsus, 2007; Adewumi & Ige, 2021).

The Open University Learning Analytics Dataset (OULAD) is a widely used benchmark for research in learning analytics due to its large scale and real-world nature (Kuzilek

et al., 2017). Numerous studies have leveraged the OULAD dataset to address various research questions. For example, some have focused on predicting student performance and dropout rates using machine learning models (Khalil & Ebner, 2016; O'Grady et al., 2021), while others have explored student engagement and self-regulated learning (Feki et al., 2022). Its rich clickstream data makes it a suitable testbed for developing and evaluating personalised learning recommender systems.

3.0 Methodology

We implemented a personalised recommender system using a collaborative filtering approach, following a structured methodology.

3.1 Dataset

The study utilised the Open University Learning Analytics Dataset (OULAD). The dataset provided by the UK's Open University contains student interaction logs, demographic information, and course metadata. It's a valuable resource for studying student engagement and developing educational technologies due to its large scale and real-world data. The dataset is a suitable testbed for developing and evaluating personalised learning recommender systems.

The dataset comprised 26,074 learners and 22 courses. We focused specifically on the `studentVle.csv` and `courses.csv` tables to capture learner-course interactions. The interaction data, primarily clickstream activity, was used to build the user-item matrix. As is typical with educational datasets, this matrix was highly sparse, with approximately 95% of its entries being zero, indicating a lack of interaction for the vast majority of learner-course pairs. The distribution of interactions was highly skewed, with a few courses attracting a disproportionately high number of enrollments, as shown in Figure 1. This skew is a common characteristic of educational data, explaining

why a simple popularity-based recommender can sometimes perform reasonably well (Kuzilek et al., 2017).

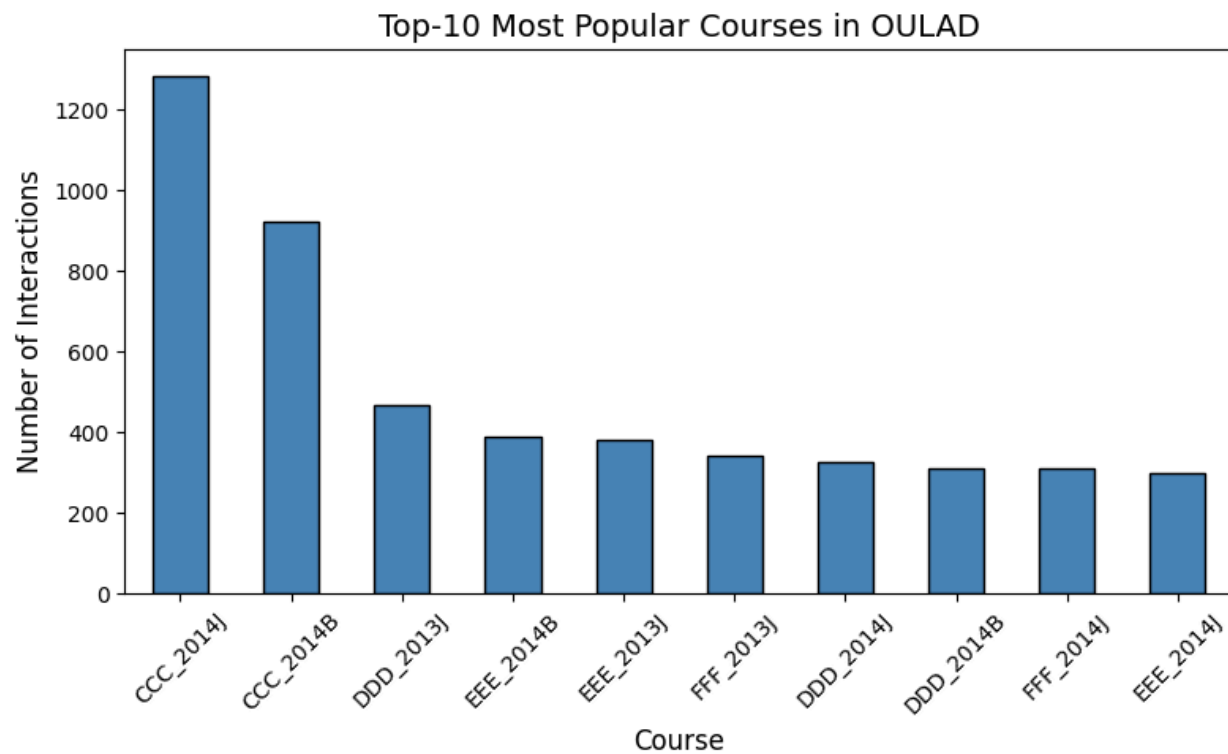


Figure 1: Top 10 most popular courses in OULAD

The figure illustrates the skewed distribution of learner interactions, with a small number of courses attracting the majority of enrollments.

3.2 Preprocessing

- **Data Aggregation:** We aggregated the clickstream activity for each learner-course pair.
- **Rating Normalisation:** The aggregated click data was normalised into a 1-5 rating scale.
- **User Filtering:** To support the leave-one-out evaluation protocol, we filtered out users with fewer than two interactions.

3.3 Model

We implemented a Singular Value Decomposition (SVD) collaborative filtering algorithm, a matrix factorisation technique that learns latent factors for both users and items (Koren et al., 2009).

3.4 Evaluation Metrics

To evaluate the performance of our recommender, we used three standard metrics:

- Root Mean Squared Error (RMSE): Measures the accuracy of rating predictions.
- Hit Rate (HR@K): Measures whether the actual enrolled course is present in the top-K recommended list.
- Normalised Discounted Cumulative Gain (NDCG@K): Rewards ranking the correct course higher in the top-K list.

3.5 Baseline

We compared our SVD model against a non-personalised popularity-based recommender. This baseline simply recommends the courses with the highest number of interactions, serving as a robust benchmark against which our personalised model must demonstrate a significant improvement.

4.0 Results

4.1 Prediction Accuracy

The SVD model achieved an RMSE of 0.2544.

Recommendation Quality: The SVD model consistently outperformed the popularity baseline across different K values for both Hit Rate and NDCG. The SVD model achieved

an HR@10 of 0.4798 and a NDCG@10 of 0.2193, outperforming the baseline, which had an HR@10 of 0.4590 and an NDCG@10 of 0.2100

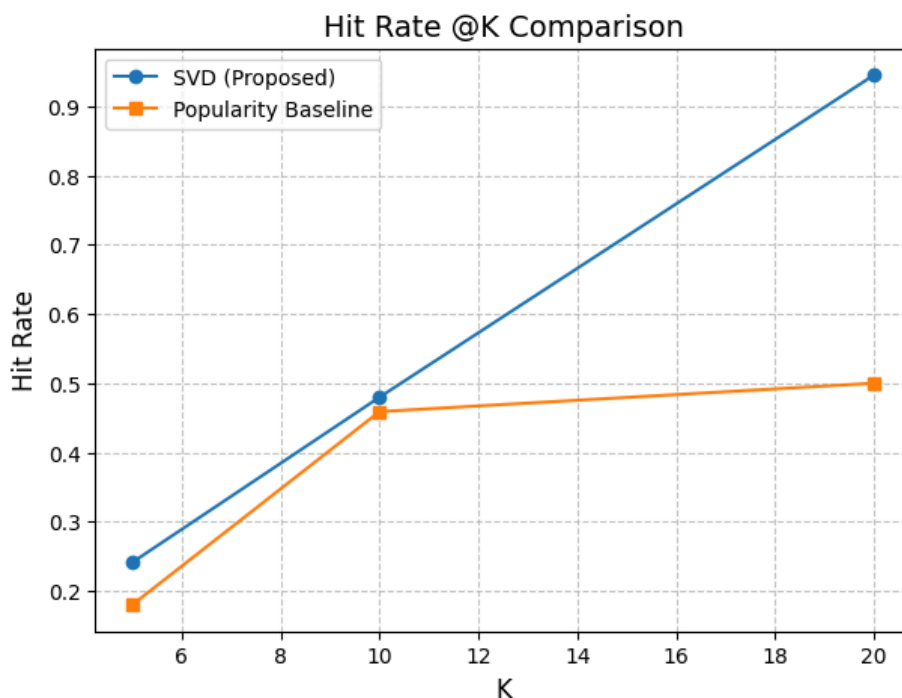


Figure 2. Hit Rate (HR@K) comparison between SVD and Popularity Baseline
The chart compares the hit rate of the SVD model against the popularity-based baseline at various K values, demonstrating the superior performance of the SVD model.

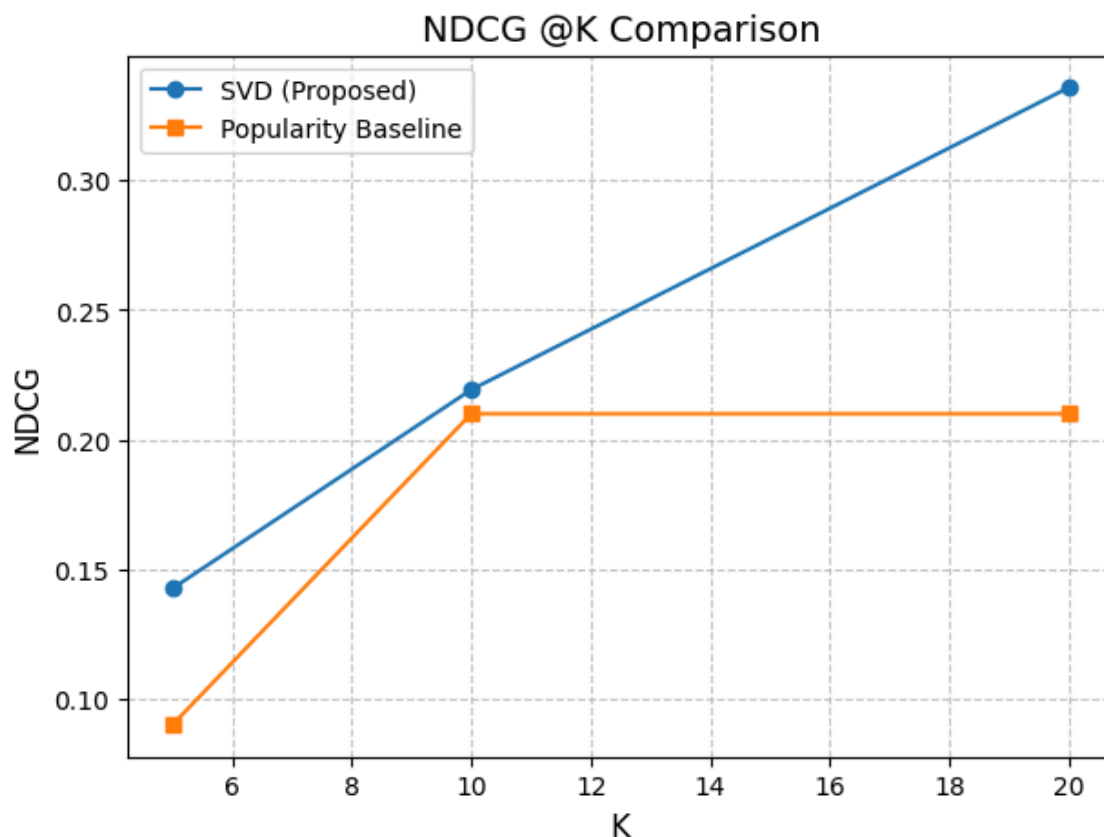


Figure 3. Normalised Discounted Cumulative Gain (NDCG@K) comparison

This figure compares the NDCG scores for the SVD model and the popularity baseline, showing how the SVD model more effectively ranks relevant courses higher in the recommendation list.

4.2 Recommendation Quality (LOO Protocol)

Model	HR@10	NDCG@10
Popularity Baseline	0.4590	0.2100
SVD (Proposed)	0.4798	0.2193

Model)

The SVD model correctly retrieved the learner's actual course in 48% of cases, outperforming the popularity baseline (46%).

5.0 Discussion

The results confirm that a collaborative filtering approach using SVD can effectively model learner-course interactions in a large-scale, sparse educational dataset. The SVD recommender consistently outperformed the popularity baseline across HR@K and NDCG@K. The improvement, while modest, highlights the potential of personalised recommenders to provide targeted guidance to learners, which is crucial for enhancing engagement in ODL (Ferguson & Clow, 2015).

Our study also highlights the limitations of a purely collaborative filtering approach. The "cold-start" problem, where recommendations cannot be made for new users or courses with no interaction data, remains a challenge. Additionally, collaborative filtering is limited by its inability to incorporate course content features, which could lead to more diverse and context-aware recommendations. Future work should explore hybrid recommenders that combine collaborative and content-based methods, as well as explainable AI approaches to improve learner trust in recommendations (Manouselis et al., 2011).

Table 1. Top-5 recommended courses for a specific learner

Learner ID	Recommended Courses	Predicted Rating
8462	FFF_2013B	1.480
	FFF_2014J	1.392
	FFF_2014B	1.356
	FFF_2013J	1.342
	AAA_2013J	1.309

This table provides a concrete example of the personalised recommendations generated by the SVD model, highlighting the differences from a generic, popularity-based list.

6.0 Conclusion

This paper demonstrated the feasibility of personalised recommenders in ODL using the OULAD dataset. By applying a collaborative filtering model based on Singular Value Decomposition (SVD), we achieved competitive performance (RMSE = 0.2544, HR@10 = 0.48, NDCG@10 = 0.22), outperforming a popularity-based baseline. These findings confirm that even classical collaborative filtering methods remain effective for modelling learner-course interactions in large-scale, sparse educational datasets. However, collaborative filtering alone has limitations, particularly in addressing sparsity and cold-start issues. While CF remains an important baseline and benchmarking tool in recommender systems research (Koren et al., 2009; Ricci et al., 2015), recent advances in hybrid and deep learning recommenders offer promising pathways for greater personalisation and robustness (Zhang et al., 2019). Future work will extend this study to hybrid recommender approaches that integrate both

interaction and content features. This would better align with the evolving needs of adaptive and explainable learning systems in ODL, ensuring that learners receive more diverse, equitable, and context-aware recommendations.

References

- Adewumi, S., & Ige, O. (2021). Hybrid recommender systems for e-learning environments: A systematic review. *Journal of Smart Environments and Green Computing*, 1(2).
- Aher, S., & Lolage, D. (2014). A review of collaborative filtering techniques in recommender systems. *International Journal of Computer Science and Information Technologies*, 5(2), 1731-1734.
- Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2012). Collaborative filtering adapted to recommender systems of e-learning. *Journal of Computer Science and Technology*, 27(6), 1109-1120. <https://doi.org/10.1007/s11390-012-1323-2>
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12(4), 331-370. <https://doi.org/10.1023/A:1021235123956>
- Feki, J., Kooli, I., & Dakhli, C. (2022). An effective approach for dealing with data sparsity and cold start problems in recommender systems. *International Journal of Information and Decision Sciences*, 14(1), 1-16. <https://doi.org/10.1504/IJIDS.2022.121520>
- Ferguson, R., & Clow, D. (2015). Learning analytics: OULAD and beyond. *Journal of Learning Analytics*, 2(3), 5-20. <https://doi.org/10.18608/jla.2015.23.2>
- Khalil, M., & Ebner, M. (2016). Learning analytics: An approach to improve learning based on student's activities. *International Journal of Emerging Technologies in Learning (iJET)*, 11(1), 17-23. <https://doi.org/10.3991/ijet.v11i01.4935>

- Khribi, M. K., Jemni, M., & Nasraoui, O. (2008). Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. In 2008 Eighth IEEE International Conference on Advanced Learning Technologies (pp. 241–245). IEEE. <https://doi.org/10.1109/ICALT.2008.198>
- Koren, Y., Bell, R., & Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8), 30–37. <https://doi.org/10.1109/MC.2009.263>
- Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open university learning analytics dataset. *Scientific Data*, 4, 170171. <https://doi.org/10.1038/sdata.2017.171>
- Manouselis, N., Drachsler, H., Verbert, K., & Duval, E. (2012). Recommender systems for learning. Springer.
- O'Grady, M. J., Mulcahy, S., & O'Hare, G. M. P. (2021). A review of recommender systems in education. *International Journal of Computer and Information Engineering*, 15(8), 564–572.
- Pazzani, M. J., & Billsus, D. (2007). Content-based recommendation systems. In *The adaptive web* (pp. 325–341). Springer.
- Ricci, F., Rokach, L., & Shapira, B. (2015). *Recommender systems handbook* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-4899-7637-6>
- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2000). Application of SVD for collaborative filtering. In *Proceedings of the 2nd International Conference on Web Information Systems Engineering* (pp. 1–10). Springer.
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1–38. <https://doi.org/10.1145/3285029>

Appendix: Code Implementation

The full code for Personalised Learning Recommender for Open and Distance Learning is available in our Google Colab notebook: [[Link](#)]

REGEX-BASED DIVIDE AND CONQUER APPROACH FOR PARSING ATM ELECTRONIC JOURNAL LOGS

Regex-Based Divide and Conquer Approach for Parsing ATM Electronic

Journal Logs

¹Arulogun, O. T., ¹Oke A. O., ²Oladipo, F. O., ¹Falohun A. S., ¹Ganiyu R. A., ³Olaniyi, O. M.,
¹Ojulari, H. O., ¹Awode R. T. and ¹Rabiu, O. E.

¹Ladoke Akintola University of Technology, Ogbomoso, ²Federal University, Lokoja,
³Federal University of Technology, Minna.

Corresponding author: otarulogun@lautech.edu.ng

Abstract

The benefits of Automated Teller Machines (ATMs) are well-documented, offering conveniences such as round-the-clock accessibility, reduced banking hall congestion, availability of cash in remote areas, multilingual support, and overall enhanced customer experience. Despite these advantages, ATMs remain highly vulnerable to a range of security threats, including card skimming, theft, vandalism, and ATM-related fraud, which significantly undermine their utility. Although anti-skimming technologies have been introduced, issues such as lost or stolen cards and shoulder surfing persist, posing substantial risks to customers and banks alike. Available studies majorly concentrated on the quantity of cash request and indiscriminate about the fraud discovery approach, without attention paid to individualised past financial activities or records. These histories, recorded in ATM Electronic Journal (EJ) logs, offer valuable insights for personalised sharp practices discovery. Contrarily, the partially systematic and unreliable/ unstable designed of EJ logs offers difficulties for required parsing system/technique. **This study** presents a programmed Regex-inclined divide-and-Conquer modality to bring out designed data from EJ logs proficiently. Established assessment on 6,469 financial activities from three important commercial banks in Nigeria presented enhanced parsing performance, with precision of 99.56%, recall of 99.93%, F1-Score of 99.74%, accuracy of 99.49%, and a misclassification rate

as low as 0.51%. These results emphasise the prospects of this proposition cum modality in improving fraud forecast structure and prognosis methodology through shopper-determined banking predictability.

Keywords: Regex, Divide and Conquer, Automated Teller Machines, ATM Electronic J

1. Introduction

The advent of Automated Teller Machines (ATMs) and similar financial deliveries has remarkably contributed to the growth of Nigeria's banking sector (Ojulari and Oke, 2016). Activities such as currency legal tender and paper money, bill transactions, financial records, inquiries, inter-bank transactions, and loan requests, previously done manually, are presently being comfortably excellently executed through ATMs. This systematisation affords financial institutions to improve and raise their financial gains, while advancing and meeting customer needs. On the other hand, ATM functions across the globe have been affected by different forms of sharp practices and cheating, which involve creating card duplicity through genetically identical copy, the use of card reader devices, and card stealing. These ill-intention practices present a critical danger to the duo financial institutions, shoppers and the individuals, as the monetary forfeitures experienced stand inevitably hard to prevent. As a result, financial institution managers persistently look for inventive ways to advance shoppers' value via advanced ATM reliability, and brilliant services proficiency. Preferably and relatively, instead of doing away with ATM, there is an increasing desire for an advanced quick-witted ATM, with the potentials to explain shoppers' financial behaviours, like the withdrawal frequency to create enterprising security awareness.

Among the critical data techniques that allows the aforesaid problem-solving is the ATM Electronic Journal (EJ), which logs detailed records of customer transactions. These EJ logs serve as vital data repositories for banks, supporting efforts to anticipate customer needs and tailor services accordingly. Despite their importance, EJ logs often lack a standardized format (Ojulari *et al.*, 2017), complicating the extraction of meaningful data for purposes such as auditing, reconciliation, and analytics. To address this challenge, this paper presents a technical framework for efficiently parsing EJ logs. The proposed method employs a divide-and-conquer strategy, combined with regular expressions, to transform unstructured log data into structured, actionable information.

2. Literature Review

The sudden increase in sharp practices associated with Automated Teller Machine (ATM) purchases has increased academic endeavours concerning developed brilliant modalities for financial activities observation and sharp practices discovery. Among the different avenues being taken advantage of, to defraud, ATMs stand out as the majorly utilised, with precedents and related references advancing in geometric frequency (Amanze and Onukwugha, 2018; Paramjit *et al.*, 2018). Accordingly, various findings have made attempts to embrace the usage of machine enabling systems, biometric and figures, to forecast, discover, or identify ATM associated sharp practices.

Brentnall *et al.* (2008) invented a statistical prototype to methodically examine time-related frequency in personalised ATM financial disposition using figures from 1,000 shoppers accounts at a UK financial institution. Subsequent study by Brentnall *et al.* (2010) embraced a multinomial distribution model to forecast personalised

withdrawal attitude during a sole ATM patronage. This model hinged on 5,000 shoppers accounts, prop-up decision-making in shopper risk administration and ATM fill-up. Venkatesh *et al.* (2014) presented a grouping system combined with computational model to predict cash request at ATM shops. By assemblage of ATMs with related withdrawal frequency and using four computational models, they identified the General Regression Neural Network (GRNN) as the important efficient way for currency request forecast. Rajwani *et al.* (2017) utilised varied methods, expanding a predictive model deploying regression and Long Short-Term Memory (LSTM) networks to explain efficient ATM withdrawals across a 2.5-year interval.

Despite the fact that these findings played remarkable roles in ATM data systematic computations, majority concentrated on summed transaction frequency or cash flow prediction rather than personalised sharp practises discovery. Furthermore, the dependence on a monitored machine learning system, demanding labelled datasets, presents difficulty in circumstances with limited categorised data. The biometric-based methodologies (e.g., Abiodun *et al.*, 2019; Ismaila *et al.*, 2018; Kavita, 2016) have also been used, but frequently entail high configuration and computational financial implications. Regardless of the shortcomings, there has been a heightened interest in using ATM Electronic Journal (EJ) logs, which accommodates required shopper transaction histories. EJ logs present prospect for instant sharp practices discovery and behaviour forecast, but are usually packed in an undefined structure with the aim of human explanation rather than machine performed. Similarly, within the associated domains like the generated files and audit trails, various parsing modalities have been invented to change the form of unadorned logs into meaningful data arrangement. Regular expression (Regex)-based parsing stands as a majorly embraced method primarily because of its speed, accuracy, and flexibility (Xu *et al.*,

2009). The use of identified capture cluster within Regex increases the readability and diagnoses of technical problems of parsing scripts (Wang and Wang, 2015), have them specifically ideal for complicated, partially defined logs.

Rombach and Fettke, (2024) carried out a study on Information Extraction (IE) from business documents using a deep learning approach, though challenges remain with generalization across different document formats. This study highlighted the importance of hybrid approaches that combine deep learning with domain-specific knowledge to ensure reliable performance in real-world scenario. Notwithstanding the credibility of Regex analysis in other distinct subsets, few studies exist on its implementation to ATM EJ logs, which are frequently incompatible arranged. These intervals showcase the need for distinct subsets solutions that can competently derive designed data from such logs. Appropriately, this study presents a Regex-based divide and Conquer framework, programmed to operate the hybrid characteristics of ATM EJ logs. This approach aims to enable efficient parsing, facilitate transaction pattern analysis, and support intelligent fraud detection strategies tailored to individual customer behaviours.

3. Methodology

The methodology for executing this work involves data acquisition, the design of the Regular expression Divide-and-conquer paradigm and the implementation

3.1 Data acquisition

Dataset of the behavioural patterns of individual ATM customer transactions was acquired from ATM Electronic Journals (EJs) of three selected Nigerian banks (NigA, NigB and NigC). This represents mixed-effects multinomial model distribution (Ojulari,

2017), consisting of 60 log files in total. The dataset comprises unstructured data features such as Date, Time, TerminalID, PAN, RRN, TSN, Transaction, Amount, Surcharge, FromAccount, ToAccount, Ledger Bal, AvailBal, and Comments. Samples of the EJs are shown in Figures 1 to 3. Table 1 shows the distribution of ATM transaction types identified in the 60 log files. Based on the transaction data count of 6,469 as shown in Table 1, the dataset was split into a 60–40 ratio. Forty (40) log files were tactically selected through a manual visual process to develop and analyse the regex patterns. This selection yielded a total of 3,848 transactions (approximately 60%) across the ATM vendors of the selected banks.

```
[2025/03/16 11:23:06] TRANSACTION DATA
[TRANSACTION RECORD]
OPCode      [ACID ]
Function ID  [5]
Card Number  [519911XXXXXX1460]
Amount      [000000000000]
Trans SEQ Number [9070]

[2025/03/16 11:23:16] EMV CANDIDATES: 1
[2025/03/16 11:23:20] EMV FINAL APP SELECTION SUCCESS: A0000000041010ebit Mastercard)
[2025/03/16 11:23:28] PIN ENTERED
[2025/03/16 11:23:37] EMV 1ST GENAC: ARQC
[2025/03/16 11:23:37] TRANSACTION REQUESTING:OPCode[ACDDAD ]
[2025/03/16 11:23:38] TRANSACTION REPLIED:FID[2] NEXT[100]
[2025/03/16 11:23:38] EMV EXTERNAL AUTHENTICATE: SUCCESS
[2025/03/16 11:23:38] EMV 2ND GENAC: TC
[2025/03/16 11:23:39] EMV TVR: 8000048000, TSI: 6800
[2025/03/16 11:23:39] REPLIED DISP COUNT IS 00002000
[2025/03/16 11:23:58] Dispense Complete
[2025/03/16 11:23:58] CASSETTE STATUS CHANGED: [OK],[LOW],[MISSING],[MISSING],[MISSING]
[2025/03/16 11:24:05] Present Completes
[2025/03/16 11:24:05] SOLICITED STATUS SEND: SUCCEEDED
[2025/03/16 11:24:07] Items Taken
[2025/03/16 11:24:08] 16\03\25 11:23 10331434
519911*****1460 009428550772 9071 009071

WITHDRAW N20 000.00
FROM ACCT NO: 234xxxxxxx

-----

SAVE TIME, AVOID TRAFFIC;DO BANKING ON THE MOVE WITH UMOBILE SMS BANKING.

THE EASE OF A VIRTUAL BANK AND TRANSACTION ALERTS
THANK YOU FOR BANKING WITH US<PostilionTran>519911*****1460|01|10|00|000002000000|20 00000|00|00|20250316|112338|20250316|009428550772|009071|00000000</PostilionTran>
[2025/03/16 11:24:08] TRANSACTION DATA
[TRANSACTION RECORD]
OPCode      [ACDDAD ]
Function ID  [2]
Card Number  [519911XXXXXX1460]
```

Figure 1: EJ of NigA

```

)1 -----
*756*03/16/2025*08:06*
CASH ADDED
TYPE 1 = 0 TYPE 2 = 0
TYPE 3 = 100 TYPE 4 = 0
* *1*R*09
*757*03/16/2025*08:06*
SUPERVISOR MODE EXIT
* *1*P*20,M-
*758*03/16/2025*08:08*
*1*y*0,M-980000000000000,R-0,S-1
*759*03/16/2025*08:08*
*PRIMARY CARD READER ACTIVATED*
*760*03/16/2025*11:21*
*TRANSACTION START*
CARD INSERTED
11:21:13 OPCODE = ACID
16\03\25 11:21 10330024
519911.....2956 9810 CARDEXPIRYALERT
???-----<PostilionTran>519911*****2956|31|40|00|000000000000|000000000000|01|01|20250316|112115|20250316|000000000000|009810|00000000</PostilionTran>
11:21:23 ATR RECEIVED T=0
11:21:33 PIN ENTERED
11:21:38 OPCODE = ACDDAD
11:21:38 GENAC 1 : ARQC
11:21:40 GENAC 2 : TC
11:21:51 NOTES STACKED
16\03\25 11:21 10330024
519911*****2956 009428542468 9811 009811
WITHDRAW N10 000.00
FROM ACCT NO: 205XXXXXX|
-----
SAVE TIME, AVOID TRAFFIC;DO BANKING ON THE MOVE WITH UMOBILE SMS BANKING.
THE EASE OF A VIRTUAL BANK AND TRANSACTION ALERTS
THANK YOU FOR BANKING WITH US<PostilionTran>519911*****2956|01|10|00|000001000000|10 00000|00|00|20250316|112139|20250316|009428542468|009811|00000000</PostilionTran>
11:21:55 NOTES PRESENTED 0,0,10,0

CASH TOTAL TYPE1 TYPE2 TYPE3 TYPE4
DENOMINATION 500 200 1000 100
DISPENSED 00000 00000 01798 00000

```

Figure 2: EJ of NigB

```

08:20:17 TRANSACTION REQUEST ACID
08:20:17 TRANSACTION REPLY NEXT 018 FUNCTION 5000
16\03\25 08:21 10331863
519911.....8979 9629 CARDEXPIRYALERT
???-----<PostilionTran>519911*****8979|31|40|00|000000000000|000000000000|01|01|20250316|082114|20250316|000000000000|009629|00000000</PostilionTran>
08:20:19 EMV CANDIDATES: 1
08:20:21 EMV AID A0000000041010 / 519911*****8979 STARTED
08:20:29 PIN ENTERED
08:20:54 AMOUNT 1900000 ENTERED
08:20:55 TRANSACTION REQUEST ACAAAD
08:20:55 INTERACTIVE TRANSACTION REPLY
08:20:57 EXTERNAL AUTHENTICATE: NO ARPC
08:20:57 TRANSACTION REPLY NEXT 092 FUNCTION 5000
08:20:58 TVR: 8000040000, TSI: 6000
08:21:06 EMV AID A0000000041010 / 519911*****8979 STARTED
08:21:22 PIN ENTERED
08:21:43 AMOUNT 1900000 ENTERED
08:21:43 TRANSACTION REQUEST ACDDAD
08:21:44 TRANSACTION REPLY NEXT 100 FUNCTION 2086
08:21:44 TVR: 8000040000, TSI: 6000
08:21:51 CASH REQUEST: 00001900
08:21:51 CASH 3:2,19;
08:21:56 CASH PRESENTED
16\03\25 08:22 10331863
519911*****8979 009427837265 9631 009631
LARGE SUM REQUESTED =
WITHDRAW N19 000.00
FROM ACCT NO: 232XXXXXX
-----
SAVE TIME, AVOID TRAFFIC;DO BANKING ON THE MOVE WITH UMOBILE SMS BANKING.
THE EASE OF A VIRTUAL BANK AND TRANSACTION ALERTS
THANK YOU FOR BANKING WITH US<PostilionTran>519911*****8979|01|10|00|000001900000|19 00000|00|00|20250316|082241|20250316|009427837265|009631|00000000</PostilionTran>
08:22:00 CASH TAKEN
08:22:20 CARD(519911*****8979) TAKEN
08:22:22 <- TRANSACTION END
*250*08:24:13 SAFE DOOR OPENED
*251*08:24:17 SECOND CASSETTE REMOVED
*252*08:25:02 SECOND CASSETTE INSERTED
*253*08:25:11 SERVICEMODE ENTERED
CLEAR CASH

```


Figure 3: EJ of NigC

Table 1: Transaction distribution across 3 ATM vendors

Transaction Events	Frequency in the EJ
MOBILE AIRTIME TOP-UP	132
CHANGE PIN	78
INQUIRY	1,200
INTERBANK TRANSFER	731
UNKNOWN*	378
WITHDRAW	3,950
Total Transaction Count:	6,469

*The unknown transactions are typically failed transactions that do not resolve at the host server.

3.2 Divide and Conquer Paradigm

Divide and Conquer Paradigm is a powerful problem-solving technique that works by breaking a problem into smaller, more manageable subproblems, solving each subproblem independently, and then combining those solutions to solve the original problem. The parsing approach begins by dividing the EJ log stream into individual transaction sessions using identifiable delimiters (transaction start and transaction end). Each session is then conquered by applying a comprehensive regex pattern capable of extracting multiple transactional fields. This segmentation improves performance and avoids excessive backtracking common with applying regex to large unbounded text. Algorithm 1 shows the process of the EJ stream using divide and conquer paradigm.

Input:

EJ file stream (FileStream)

Session delimiters: {transaction start, transaction end}

Regex pattern to extract transaction fields

Output:

Data array of extracted transaction fields

Algorithm 1: Process EJ File Stream Using Divide and Conquer

BEGIN

// Step 0: Read the EJ file as a string

FileStream ← Open EJ file

ejstream ← FileStream.Read() as string

// Step 1: Split EJ stream into transaction sessions using delimiters

session_delimiters ← {transaction_start_marker, transaction_end_marker}

ejsessionArray ← Split(ejstream, session_delimiters)

data ← empty array

i ← 0

// Step 2: Process each transaction session

WHILE i < Length(ejsessionArray) DO

ejsession ← ejsessionArray[i]

```
// Step 3: Apply regex to extract transaction fields
```

```
matchArray ← Match(ejsession, regex_pattern)
```

```
IF matchArray IS NOT null THEN
```

```
    // Step 4: Tokenize and extract fields
```

```
    pan ← Tokenize(matchArray, "pan")
```

```
    tsn ← Tokenize(matchArray, "tsn")
```

```
    date ← Tokenize(matchArray, "date")
```

```
    time ← Tokenize(matchArray, "time")
```

```
    terminal ← Tokenize(matchArray, "terminal")
```

```
    transtype ← Tokenize(matchArray, "transtype")
```

```
    currency ← Tokenize(matchArray, "currency")
```

```
    amount ← Tokenize(matchArray, "amount")
```

```
    account ← Tokenize(matchArray, "account")
```

```
    ledger ← Tokenize(matchArray, "ledger")
```

```
    avail ← Tokenize(matchArray, "avail")
```

```
    comment ← Tokenize(matchArray, "comment")
```

```
    error ← Tokenize(matchArray, "error")
```

```
    status ← Tokenize(matchArray, "status")
```

```
// Add the extracted data to the data array
```

```
Add(data, {
```

```
    pan, tsn, date, time, terminal, transtype, currency,
```

```
    amount, account, ledger, avail, comment, error, status
```

```
})
```

```
ENDIF
```

```
// Step 5: Move to next session
```

```
i ← i + 1
```

```
ENDWHILE
```

```
RETURN data
```

```
END
```

3.3 Regular Expression (Regex) Design

Regex offers concise and expressive syntax for capturing complex text patterns that would otherwise require extensive procedural code. With powerful and precise pattern matching capabilities beyond simple string comparisons, regex is widely supported across programming languages and tools, making it a versatile and transferable skill. Its optimized engines deliver efficient performance in key tasks such as data validation, text extraction, transformation, and cleaning, making it an ideal choice for processing unstructured Electronic Journal (EJ) log data (Friedl, 2006). Familiar Regex symbols with meanings are presented as Table A of the Appendix. Within this study, attentively structured Regex pattern, employing identified capture clusters, is employed to derive required areas from EJ logs. These areas are further delineated to a designed schema, *Ej Transaction*, comprehensive in Table B of the Appendix, which safeguard the compatibility with different retention systems like similar databases or flat files (e.g., CSV). The Regex structure is conceived to accommodate different

financial types like withdrawal, inquiry deposit, and account transfer, while productively taking care of optional components and formatting requirements.

The application was conducted by Microsoft Visual Studio 2022 using C# and applied on Windows 11 x64. Patterns were processed using *RegexOptions.Compiled* to improve execution, with the following criteria: matches found, parsing time and No. of files. The validation was carried out with the Espresso application. Error detection was powerfully-built, logging distorted lines and authorize iterative clarification of the Regex. To prevent execution challenges, Regex was implemented purposely to pre-divide session blocks, thus averting fatal backtracking and enhancing runtime performance. Algorithm 2 presents the Regex pattern applied to derive fields from EJ logs of the three financial institutions, assembled into five divisions: *Transaction Details Extraction; Transaction Type, Currency, and Amount Recognition; Account and Surcharge Information Parsing; Network and Account Balance Extraction; and Multi-line Comments Capture*.

Algorithm 2

```
(?:\[\d{4}\]\[\d{2}\]\[\d{2}\]\s\d{2}:\d{2}:\d{2}\])?\s*(?<Date>\d{2}[-\/\]\d{2}[-\/\]\d{2})\s+(?<Time>\d{2}:\d{2})\s+(?<TerminalID>\w{8,11})\s*(?:\r*\n*)*
[-]*\s*(?:\r*\n*)*
(?:CARD\s?NUMBER)?\s*(?<PAN>\d*[Xx.]+\d+)\s*(?:\r*\n*)
((?:SEQ)?\s*(?:NUMB)?\s*)?
#(?:(?<RRN>\d{10,12}I[?]+)?\s*(?:\r*\n*)*(?<TSN>\d{4})I(?<TSN>\d{4})\s*(?:\r*\n*)
* (?<RRN>\d{10,12}I[?]+)?\s*(?<otherNum>\d+I[?]+)?(?:\r*\n*)*
((?<RRN>\d{10,12}I[?]+)?\s*(?<TSN>\d{4})I(?<TSN>\d{4})\s*(?<RRN>\d{10,12}I[?]+)
?)\s*(?<otherNum>\d+I[?]+)?(?:\r*\n*)*
((?:AUTH\s*NUM)?\s*(\w*)?\s*(?:\r*\n*)*)?
(?:LARGE\s+SUM\s+REQUESTED\s*=)?(?:\r*\n*)*
```

```
(((<Transaction>WITHDRAW)\s+(?<CurrencyCode>Nl\w{3})?(?<Amount>\d+(?:\s*\d{3})*\.\d{2}))?|
((?<Transaction>MINISTATEMENT)\s*)|
((?<Transaction>THIRD\s+PARTY\s+PAYMENT)\s+(?<CurrencyCode>Nl\w{3})?(?<Amount>\d+(?:\s*\d{3})*\.\d{2}))?|
((?<Transaction>INTERBANK\s+TRANSFER[TRANSFER])\s+(?<CurrencyCode>Nl\w{3})?(?<Amount>\d+(?:\s*\d{3})*\.\d{2}))?|
((?<Transaction>INQUIRY))|
((?<Transaction>ADVANCE\s+PREPAID)\s+(?<CurrencyCode>Nl\w{3})?(?<Amount>\d+(?:\s*\d{3})*\.\d{2}))?|
((?<Transaction>DEPOSITBNA\s+DEPOSITPAYMENT\s+BY\s+BNA\s+DEPOSIT)\s+(?<CurrencyCode>Nl\w{3})?(?<Amount>\d+(?:\s*\d{3})*\.\d{2}))?|
((?<Transaction>(\w+\s*\w+))\s+((?<CurrencyCode>Nl\w{3})?(?<Amount>\d+(?:\s*\d{3})*\.\d{2}))?)?|
((?<Transaction>CHANGE\s+PIN)?\s+))?\r*
(?:FROM[:]\s+(?:ACCT\s+NO[:]\s*)?(?<FromAccount>[\w.]+))?\s*\r*
(?:TO[:]\s+(?:ACCT\s+NO[:]\s*)?(?<ToAccount>[\w.]+))?\s*(\r*\n*)*
(SURCHARGE[:]\s+(?<CurrencyCode>Nl\w{3})?(?<Surcharge>\d+(?:\s*\d{3})*\.\d{2}))?\s*\r*\s*
(?<Network>(NETWORK[:]\s+(\w+.*(\r*\n*)+)*\r*\n*\s*\w+.*))?\s*\r*\s*
(((?<Ledger>(LEDGER))\s+(?<CurrencyCode>Nl\w{3})?(?<LedgBal>\d+(?:\s*\d{3})*\.\d{2}))?\r*)
(((?<Avail>(AVAIL))\s+(?<CurrencyCode>Nl\w{1,3})?(?<AvailBal>\d+\r?\n?\w*(?:\s*\d{3})*\.\d{2}))?\r*)
(?<Comments>\s*\w+.*\r*\n*\w+.*)?
```

3.3.1 Transaction Details Extraction

This Regex is designed to parse Electronic Journal logs, deriving important characteristics like the optional timestamp header, a date and time in various arrangements, a terminal identifier, and a concealed card number (PAN) identified as "CARD NUMBER." It equally expresses Reference Numbers (RRN), Transaction

Sequence Numbers (TSN), and added numerical fields, irrespective of their arrangement or frequency. It operates voluntary labels like: "SEQ," "NUMB," "AUTH NUM," and "LARGE SUM REQUESTED," and is pliable about whitespace, newlines, and dividers, ensuring that it is sturdy for different acceptable arrangements.

3.3.2 Transaction Type, Currency, and Amount Recognition

This Regex recognizes extensive scope of banking transaction forms like: "MOBILE AIRTIME TOP-UP," "CHANGE PIN," "INQUIRY," "INTERBANK TRANSFER," "UNKNOWN," and "WITHDRAW," as shown in Table 1. For financial activities that entail money, it offers the currency code (a single letter like "N" or three-letter codes) and the specific value with agreement for alternative thousand dividers and denominators. The identified clusters simplify separation of transaction type, currency, and amount, modifying to different intervals and alternative fields.

3.3.3 Account and Surcharge Information Parsing

This pattern derives alternative "FROM" and "TO" bank details, to afford labels like "ACCT NO" and showing account identifiers consist of word characters and dots. Besides, it alternatively shows a surcharge amount, with the currency code and monetary value, and thousand separators with decimals, following the label "SURCHARGE." The regex operates flexible whitespace, and line breaks to safely parse sender and receiver account details with surcharge charges from transaction histories.

3.3.4 Network and Account Balance Extraction

The Regex alternatively encapsulates multi-line network detail starting with "NETWORK," circumscribe subsequent lines that give information. It closely recognizes

ledger balances identified as 'LEDGER' and obtainable balances classified as 'AVAIL,' capturing related currency codes and monetary quantities arranged with alternative thousand separators and mandatory digits. Its liberality for whitespace and line breaks affords it suitability for deriving organized network and financial statements that reflect balance data.

3.3.5 Multi-line Comments Capture

This Regex recourse matches one or two lines of freeform text under an identified cluster comments, accepting leading whitespace and newlines. It captures blocks of descriptive or comment text that span one or two lines, useful for extracting brief textual notes or remarks from a document or record.

3.4 Performance Evaluation

The performance of the approach was evaluated using the following metrics: precision, recall, accuracy, and F1-score. The pseudo-code to calculate the performance metrics is as shown in Algorithm 3 and shown in equations 1 to 4. Regex evaluation is often manual, especially during the development or modification of regex patterns, because it is highly contextual and the ground truth can be ambiguous.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \quad (1)$$

$$\text{Recall (Sensitivity)} = \text{TP} / (\text{TP} + \text{FN}) \quad (2)$$

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \quad (3)$$

$$\text{F1-score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (4)$$

Where, TP = True Positive, shows correctly predicted positive cases.

TN = True Negative, shows correctly predicted negative cases.

FP = False Positive, shows predicted positive, but actually negative (*Type I error*).

FN = False Negative, shows predicted negative, but actually positive (*Type II error*).

Precision indicates how many predicted positives are truly positive, Recall indicates how many actual positives a

Algorithm 3: Pseudo-code to calculate performance metrics

```

int TP = 0, TN = 0, FP = 0, FN = 0;

foreach (var (input, expected) in testCases)
{
    bool actual = regex.IsMatch(input);

    if (actual && expected) TP++;
    else if (!actual && !expected) TN++;
    else if (actual && !expected) FP++;
    else if (!actual && expected) FN++;
}

double accuracy = (double) (TP + TN) / (TP + TN + FP + FN);
double precision = TP + FP == 0 ? 0 : (double)TP / (TP + FP);
double recall = TP + FN == 0 ? 0 : (double)TP / (TP + FN);
double f1Score = (precision + recall) == 0 ? 0 : 2 * precision * recall / (precision + recall);

Console.WriteLine($"Accuracy: {accuracy:P2}");
Console.WriteLine($"Precision: {precision:P2}");
Console.WriteLine($"Recall: {recall:P2}");
Console.WriteLine($"F1 Score: {f1Score:P2}");

```

4. Results and Discussion

The outcomes of applying a Regex-Based Divide and Conquer approach to parse and extract customer transaction data from ATM Electronic Journal (EJ) logs of three different ATM vendors: Hyosung, NCR, and Wincor, across NigA, NigB, and NigC, respectively, are presented. Parsing was successful for most session blocks, while failed matches were logged for further analysis. Figures 4, 5, and 6 show sample structured log extractions and each Figure presents an extract of parsed records in a tabular form, breaking down various attributes of ATM transactions. While all three vendor logs yielded similar data fields, the parsing performance, match count, and transaction characteristics differed significantly, showcasing both the parser's versatility and the distinct nature of each vendor's log structure. Across all three vendors, the parser consistently extracted a standardized set of core fields, including: Date, Time, TerminalID, PAN, RRN, TSN, Transaction, CurrencyCode, Amount, Surcharge, FromAccount, ToAccount, Ledger Bal, AvailBal and Comment. This common structure enables uniform analysis, regardless of ATM brand, but differences in parsing outcomes reveal vendor-specific patterns and complexities.

4.1 NigA Result

The NigA logs produced 1,616 matches from eleven (11) log files (randomly selected) in just 16.99 seconds as shown in Figure 4. It has the fastest parsing time among the three vendors for the subset of the forty (40) logs. The dataset includes a variety of transaction types such as withdraw, inquiry, deposit, transfer, recharge, and unknown. Uniquely, it contains many WITHDRAW (ON US and OFF US) transactions (N1000–N20,000), which are relatively common in standard ATM operations and likely

related to cash withdrawal. The "Comment" field features meaningful indicators such as "TRANSACTION COMPLETED," "CARDEXPIRYALERT," and "NOT ENOUGH FUNDS," highlighting the parser's ability to detect both successful and failed transactions. The "PAN" entries are mostly masked to preserve confidentiality. Overall, HYOSUNG's result demonstrates quick, diverse, and efficient parsing.

ATM Vendor

Journal File

HYOSUNG

D:\r... \EJs\Hyosung

Load File

Load Folder

Parse

Save

11 files: parsing in progress...
1616 Matches; Parsing time: 00:00:16.99

TerminalID	TransDate	TSN/STAN	PAN	TransTypeid	TransType	TransCat	Currency	Amount	Surcharge	AvailBalance	Ledger	AccountFrom	AccountTo	R-R-N	Detail
10331434	3/16/2025 11:23:00 AM	9070	519911*****1460	1	UNKNOWN	ON US	N	0	0	0	0				TRANSACTION. ^
10331434	3/16/2025 11:23:00 AM	9071	519911*****1460	2	WITHDRAW	ON US	N	20000	0	0	0	2344124767		009428550772	TRANSACTION.
10331434	3/16/2025 11:24:00 AM	9072	519911*****1460	2	WITHDRAW	ON US	N	20000	0	2025	0	2344124767		009428554806	TRANSACTION.
10331434	3/16/2025 11:25:00 AM	9073	519911*****1460	2	WITHDRAW	ON US	N	20000	0	0	0	2344124767		009428558361	TRANSACTION.
10331434	3/16/2025 11:25:00 AM	9074	519911*****1460	1	UNKNOWN	ON US	N	20000	0	0	0				TRANSACTION.
10331434	3/16/2025 11:32:00 AM	9075	519911*****6461	1	UNKNOWN	ON US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 11:33:00 AM	9076	519911*****6461	1	UNKNOWN	ON US	N	1000	0	0	0				TRANSACTION.
10331434	3/16/2025 11:33:00 AM	9077	519911*****6461	1	UNKNOWN	ON US	N	1000	0	0	0				TRANSACTION.
10331434	3/16/2025 2:34:00 PM	9078	519911*****9480	1	UNKNOWN	ON US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 2:34:00 PM	9079	519911*****9480	1	UNKNOWN	ON US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 2:34:00 PM	9080	519911*****9480	2	WITHDRAW	ON US	N	20000	0	2025	0	2027257674		009429404967	TRANSACTION.
10331434	3/16/2025 2:35:00 PM	9081	519911*****9480	2	WITHDRAW	ON US	N	20000	0	0	0	2027257674		009429407830	TRANSACTION.
10331434	3/16/2025 4:08:00 PM	9082	539983*****0862	1	UNKNOWN	OFF US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 4:09:00 PM	9083	539983*****0862	2	WITHDRAW	OFF US	N	10000.00	107.50	0	0	0009750172		009429785800	TRANSACTION.
10331434	3/16/2025 5:03:00 PM	9084	539983*****9506	1	UNKNOWN	OFF US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 5:04:00 PM	9085	539983*****9506	2	WITHDRAW	OFF US	N	5000.00	107.50	0	0	0177360670		009430042917	TRANSACTION.
10331434	3/16/2025 5:04:00 PM	9086	519911*****9480	1	UNKNOWN	ON US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 5:05:00 PM	9087	519911*****9480	2	WITHDRAW	ON US	N	20000	0	2025	0	2027257674		009430049042	TRANSACTION.
10331434	3/16/2025 5:08:00 PM	9088	519911*****7659	1	UNKNOWN	ON US	N	0	0	0	0				TRANSACTION.
10331434	3/16/2025 5:08:00 PM	9089	519911.....7659	1	UNKNOWN	ON US	N	20000							TRANSACTION. v

Figure 4: Sample structured log extraction from Hyosung, NigA.

4.2 NigB Result

The NigB logs produced 233 matches in 26.61 seconds as shown in Figure 5, placing it between Hyosung and Wincor in parsing speed. The file contents are complex and verbose due to the associated device logs appended to the EJs. The extracted data was primarily focused on identifying the transaction subset within the verbose content. The parser also captured surcharge data for OFF US transactions, such as a

₦1300 fee. The *AccountFrom* column includes entries such as 'CHECK' for OFF-US transactions, while the *Comment* field contains indicators like 'CARDEXPIRYALERT,' 'YOU HAVE EXCEEDED YOUR WITHDRAWAL FREQUENCY LIMIT,' and 'TRANSACTION COULD BE COMPLETED.' Although the match count is relatively low due to verbose device logs appended to the EJs, the parsing time is moderate and demonstrates extraction clarity.

4.3 NigC Result

The NigC logs yielded the highest number of matches (1,999) from 19 files randomly selected as shown in Figure 6, but also had the longest parsing time of approximately 1 minute. This suggests a greater log complexity, more detailed content, and additional overhead from reading 19 files into memory. Despite some fields like "Amount" and "AccountFrom" being zero or empty, the parser extracted rich structures. The "TransType" column includes "UNKNOWN ON-US," "INQUIRY ON-US," and "WITHDRAW ON-US". The results suggest that the parser is highly optimized for WINCOR logs, especially those with rich metadata.

ATM Vendor: NCR Journal File: D:\li IEJs\NCR

Load File Load Folder Parse Save

10 files: parsing in progress...
233 Matches; Parsing time: 00:00:26.61

TerminalID	TransDate	TSN/STAN	PAN	TransType	TransType	TransCat	Currency	Amount	Surcharge	AvailBalance	Ledger	AccountFrom	AccountTo	R-R-N	Detail
12560034	3/16/2025 1:42:00 PM	9931	464978.....6971	2	WITHDRAW	OFF US	N	350000	1300	0	0			122400000022	*886*03/16/2025*01 ^
12290026	3/16/2025 10:36:00 AM	3903	521603.....7999	2	WITHDRAW	OFF US	N	10000	0	6314	0	510500050844		510500050844	OPCODE = ACCAE
12290026	3/16/2025 11:44:00 AM	3904	421758.....8658	3	INQUIRY	ON US	N	0	0	0	0	501020005698		501020005698	OPCODE = ABABE
12290026	3/16/2025 11:46:00 AM	3905	421758.....4793	2	WITHDRAW	ON US	N	150000	0	842717	0	502010022237		502010022237	OPCODE = ACCAE
10330024	3/16/2025 11:21:00 AM	9810	519911.....2956	1	UNKNOWN	ON US	N	0	0	0	0				OPCODE = ACID
10330024	3/16/2025 11:21:00 AM	9811	519911.....2956	2	WITHDRAW	ON US	N	10000.00	0	0	0	2053777085		009428542468	OPCODE = ACDD/
10330024	3/16/2025 11:22:00 AM	9812	519911.....2956	2	WITHDRAW	ON US	N	20000.00	0	0	0	2053777085		009428545074	OPCODE = ACDD/
10330024	3/16/2025 11:22:00 AM	9813	519911.....2956	2	WITHDRAW	ON US	N	10000.00	0	0	0	2053777085		009428547732	OPCODE = ACDD/
10330024	3/16/2025 11:31:00 AM	9814	506102.....4919	1	UNKNOWN	ON US	N	0	0	0	0				OPCODE = ACID
10330024	3/16/2025 11:31:00 AM	9815	506102.....4919	2	WITHDRAW	ON US	N	20000.00	0	0	0	2021493245		009428594683	OPCODE = ACDD/
10330024	3/16/2025 11:32:00 AM	9816	506102.....4919	2	WITHDRAW	ON US	N	20000.00	0	0	0	2021493245		009428587272	OPCODE = ACDD/
10330024	3/16/2025 11:32:00 AM	9817	506102.....4919	2	WITHDRAW	ON US	N	20000.00	0	0	0	2021493245		009428590209	OPCODE = ACDD/
12320010	3/14/2025 2:27:00 PM	1295	471442.....4651	3	INQUIRY	ON US	N	0	0	629.02	0	540320750033183		009420227356	OPCODE = ABABE
12560028	3/16/2025 11:14:00 AM	7447	460044.....9456	2	WITHDRAW	ON US	N	100000	0	78840	0				*876*03/16/2025*01
12290025	3/16/2025 7:19:00 AM	5872	457793.....1977	2	WITHDRAW	ON US	N	9000	0	0	0	CHECK		009427633350	OPCODE = ACCAE
12290025	3/16/2025 9:39:00 AM	5873	421758.....6309	2	WITHDRAW	ON US	N	1000	0	5550	0	503010022494		503010022494	OPCODE = ACCAE
12290025	3/16/2025 9:40:00 AM	5874	421758.....6309	2	WITHDRAW	ON US	N	1000	0	4550	0	503010022494		503010022494	OPCODE = ACCAE
12290025	3/16/2025 10:48:00 AM	5875	421758.....0539	3	INQUIRY	ON US	N	0	0	2003261	0	506500007731		506500007731	OPCODE = ABABE
12290025	3/16/2025 10:51:00 AM	5876	471336.....3608	1	UNKNOWN	OFF US	N	0	0	0	0			009428414161	OPCODE = ABABE
12290025	3/16/2025 10:52:00 AM	5877	471336.....3608	1	UNKNOWN	OFF US	N	0	0	0	0			009428418615	OPCODE = ABABE

Figure 5: Sample structured log extraction from NCR, NigB

ATM Vendor: WINCOR Journal File: D:\Inf . s\IEJs\wincor

Load File Load Folder Parse Save

19 files: parsing in progress...
1999 Matches; Parsing time: 00:01:00.84

TerminalID	TransDate	TSN/STAN	PAN	TransType	TransType	TransCat	Currency	Amount	Surcharge	AvailBalance	Ledger	AccountFrom	AccountTo	R-R-N	Detail
10331497	3/16/2025 8:07:00 AM	3247	506115.....0470	1	UNKNOWN	OFF US	N	0	0	0	0				08:06:28 TRACK ^
10331497	3/16/2025 8:24:00 AM	3248	492069.....5736	1	UNKNOWN	OFF US	N	0	0	0	0				08:24:01 TRACK
10331497	3/16/2025 8:26:00 AM	3249	492069.....5736	5	DEPOSIT	OFF US	N	240000.00	0	0	0	1012851932		009427849458	PIN ENTERED
10331497	3/16/2025 9:11:00 AM	3250	492069.....7585	1	UNKNOWN	OFF US	N	0	0	0	0				09:10:50 TRACK
10331497	3/16/2025 9:11:00 AM	3251	492069.....7585	2	WITHDRAW	OFF US	N	20000.00	0	0	0	2197980459		009428020226	PIN ENTERED
10331497	3/16/2025 9:12:00 AM	3252	492069.....7585	2	WITHDRAW	OFF US	N	20000.00	0	0	0	2197980459		009428023044	PIN ENTERED
10331497	3/16/2025 9:13:00 AM	3253	492069.....7585	2	WITHDRAW	OFF US	N	20000.00	0	0	0	2197980459		009428026179	PIN ENTERED
10331497	3/16/2025 9:13:00 AM	3254	492069.....7585	2	WITHDRAW	OFF US	N	20000.00	0	0	0	2197980459		009428028593	PIN ENTERED
10331497	3/16/2025 9:14:00 AM	3255	492069.....7585	2	WITHDRAW	OFF US	N	20000.00	0	0	0	2197980459		009428031250	PIN ENTERED
10331497	3/16/2025 9:15:00 AM	3256	492069.....8217	1	UNKNOWN	OFF US	N	0	0	0	0				09:15:00 TRACK
10331497	3/16/2025 9:16:00 AM	3257	492069.....8217	2	WITHDRAW	OFF US	N	10000.00	0	0	0	2064415503		009428038618	PIN ENTERED
10331497	3/16/2025 9:23:00 AM	3258	519911.....9144	1	UNKNOWN	ON US	N	0	0	0	0				09:22:51 TRACK
10331497	3/16/2025 9:24:00 AM	3259	519911.....9144	2	WITHDRAW	ON US	N	20000.00	0	0	0	2188088159		009428069040	PIN ENTERED
10331497	3/16/2025 9:24:00 AM	3260	519911.....9144	2	WITHDRAW	ON US	N	20000.00	0	0	0	2188088159		009428070900	PIN ENTERED
10331497	3/16/2025 9:26:00 AM	3261	519911.....9144	1	UNKNOWN	ON US	N	0	0	0	0				09:25:43 TRACK
10331497	3/16/2025 9:26:00 AM	3263	519911.....9144	1	UNKNOWN	ON US	N	0	0	0	0				09:26:10 TRACK
10331497	3/16/2025 9:27:00 AM	3264	519911.....9144	2	WITHDRAW	ON US	N	10000.00	0	0	0	2188088159		009428080910	PIN ENTERED
10331497	3/16/2025 9:37:00 AM	3265	519911.....1034	1	UNKNOWN	ON US	N	0	0	0	0				09:36:58 TRACK
10331497	3/16/2025 9:39:00 AM	3266	519911.....1034	5	DEPOSIT	ON US	N	35000.00	0	0	0	2003712911		009428130227	PIN ENTERED
10331497	3/16/2025 9:53:00 AM	3267	519911.....7585	1	UNKNOWN	ON US	N	0	0	0	0				09:52:55 TRACK

Figure 6: Sample structured log extraction from Wincor, NigC

4.4 Comparative Analysis

A comparative analysis showed that the contents of the observed log files differ (NigA=11, NigB=10, NigC=19), due to varying customer activities and fluctuations in ATM usage. Hence, ATM traffic directly influences the number of transactions recorded in each log file. NigA demonstrates superior efficiency, rapidly extracting a diverse set of transactions with good completeness and clear details, indicating a relatively straightforward log format. NigB, conversely, is the least efficient with the lowest match count despite moderate parsing time, suggesting a more complex or inconsistent log format requiring significant regex refinement for improved completeness and accuracy. NigC achieves the highest volume of accurate and detailed extractions, capturing rich metadata and card-processing trails, but at the cost of being the slowest, pointing to a highly complex and verbose log structure that necessitates extensive processing, as shown in Table 2.

Table 2. Comparative analysis of the performance criteria

Vendor	Matches found	Parsing time	No. of Files	Remarks
Hyosung (Nig A)	1,616	16.99 s	11	Relatively fast and consistent in extracting transactions from 11 EJ log files. Effective at identifying and categorizing diverse transactions
NCR (Nig B)	233	26.61 s	10	Lower matches due to low traffic on the ATMs used. The file contents are complex and verbose because of the associated device logs appended to the EJs.

Wincor (Nig C)	1,999	60.00 s	19	Highest matches from 19 files; highly effective at capturing a vast number of records and granular details, despite the time cost, indicating robust pattern matching within a min.

4.5 Evaluation Results

The results displayed previously were obtained from 40 logs across the three vendors of NigA, NigB, and NigC. The Regex-based parser was then applied to the entire set of 60 logs containing 6,469 transactions in order to evaluate its performance using precision, accuracy, recall, and F1-score. The multi-class confusion matrix shown in Table 3, illustrates how well the regex-based parser performed during extraction by comparing predicted labels with actual labels (ground truth). The table shows that the Regex-based parser is effective with 100% correctly predicted positive cases (TP), correctly predicted negative cases (TN), no Type I (FP) nor Type II errors (FN) for most field classes, including: Date, Time, TerminalID, PAN, TSN, Transaction, Amount, Surcharge, and ToAccount. However, the performance dropped on RRN, FromAccount, Ledger Bal, Avail Bal and Comment.

Table 4 illustrates the performance metrics of the regex-based parser evaluated across different transaction classes extracted from the EJ logs. These metrics provide insights into how effectively the regex-based parser identified and extracted individual classes. Table 4 shows the effectiveness of Regex-based parsing, with 100% precision, recall, accuracy, and F1-score, as well as 0% misclassification for all field classes with the exception of RRN, FromAccount, Ledger Bal, AvailBal, and Comments. This indicates that the Regex-based patterns for these fields were highly robust and consistently reliable across all test cases. The few deviations from perfect scores can

be attributed to field-specific complexities, such as variations in log format and context.

Table 3: Multi-class Confusion Matrix

Field Class	Actual Count	True Positive (TP)	True Negative (TN)	False Positive (FP)	False Negative (FN)
Date	6,469	6,469	0	0	0
Time	6,469	6,469	0	0	0
TerminalID	6,469	6,469	0	0	0
PAN	6,091	6,091	0	0	0
RRN	5,150	5,105	35	5	5
TSN	6,469	6,469	0	0	0
Transaction	6,091	6,091	0	0	0
Amount	4,813	4,813	0	0	0
Surcharge	731	731	0	0	0
FromAccount	4,680	4,550	87	23	20
ToAccount	731	731	0	0	0
Ledger Bal	5,150	5,125	25	0	0
Avail Bal	5,150	5,125	25	18	7
Comments	5,320	5,185	135	0	0

Table 4: Performance metrics

Class	Precision %	Recall %	Accuracy %	Misclassification %	F1-measure %
Date	100.00	100.00	100.00	0.00	100.00
Time	100.00	100.00	100.00	0.00	100.00
TerminalID	100.00	100.00	100.00	0.00	100.00
PAN	100.00	100.00	100.00	0.00	100.00
RRN	99.32	99.90	99.22	0.78	99.61
TSN	100.00	100.00	100.00	0.00	100.00
Transaction	100.00	100.00	100.00	0.00	100.00
Amount	100.00	100.00	100.00	0.00	100.00
Surcharge	100.00	100.00	100.00	0.00	100.00
FromAccount	98.12	99.50	97.64	2.36	98.81
ToAccount	100.00	100.00	100.00	0.00	100.00
Ledger Bal	99.51	100.00	99.51	0.49	99.76
Avail Bal	99.51	99.65	99.17	0.83	99.58
Comment	97.46	100.00	97.46	2.54	98.71

4.5.1 Average Performance

The average performance metrics (weighted metrics), derived from the multi-class confusion matrix were calculated using equations 5 to 8 and presented in Table 6, reveal that the Regex-based divide and conquer parser demonstrates outstanding performance across all evaluation measures. It achieves a precision of 99.56%, meaning it is almost always correct when extracting a class, and a recall of 99.93%,

indicating that it very rarely misses relevant cases and maintains extremely low false negatives. The F1-Score, which balances precision and recall, is equally impressive at 99.74%, confirming its strong ability to minimize both false positives and false negatives. In addition, the parser records an overall accuracy of 99.49%, highlighting its high reliability, while the misclassification rate remains exceptionally low at just 0.51%. These weighted averages underscore the consistency and robustness of the regex-based parser across all classes, even in scenarios with class imbalance. The combination of near-perfect precision, recall, and accuracy demonstrates its ability to make very few errors, making it particularly well-suited for real-world applications where both correctness and completeness are essential.

$$\text{Average Precision: } Precision = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FP_i)} \quad (5)$$

$$\text{Average Recall: } Recall = \frac{\sum_{i=1}^n TP_i}{\sum_{i=1}^n (TP_i + FN_i)} \quad (6)$$

$$\text{Accuracy: } Accuracy = \frac{\sum_{i=1}^n (TP_i + TN_i)}{\sum_{i=1}^n (TP_i + TN_i + FP_i + FN_i)} \quad (7)$$

$$\text{Average F1-Score: } F1 - Score = \frac{2 * Precision * Recall}{Precision + Recall} \quad (8)$$

Table 6: Average performance

Metric	Value	%
Average Precision	0.9956	99.56
Average Recall	0.9993	99.93
Average F1-Score	0.9974	99.74
Average Accuracy	0.9949	99.49
Average Misclassification	0.0051	0.51

5. Conclusion

This study presented a Regex-Based Divide and Conquer Approach for parsing diverse and heterogeneous ATM Electronic Journal (EJ) logs into structured data. Using a dataset of 6,469 financial activities drawn from 60 log files across three major Nigerian banks, the method efficiently extracted key standardized fields such as Date, Time, TerminalID, PAN, RRN, TSN, Transaction, Amount, Surcharge, FromAccount, ToAccount, Ledger, AvailBal, and Comments. By applying custom-tuned regex patterns adapted to vendor-specific formats, the approach demonstrated enhanced adaptability, robustness, and scalability, achieving outstanding performance with 99.56% precision, 99.93% recall, 99.74% F1-Score, 99.49% accuracy, and a misclassification rate of just 0.51%. These results underscore its reliability in handling unstructured EJ logs and its suitability for real-world financial environments. Beyond efficient parsing, the solution establishes a strong foundation for customer transaction analysis, audit trails, sharp practices discovery, fraud forecast structure, and shopper-determined banking predictability. Future extensions may include integrating machine learning for

anomaly detection and developing real-time monitoring dashboards to further strengthen banking security and performance evaluation.

6. Acknowledgement

The authors gratefully acknowledge the support of the Tertiary Education Trust Fund (TETFund), Nigeria, for providing the research grant from which this study was extracted. This work was carried out under the TETFund National Research Fund (NRF) intervention. We sincerely appreciate the management of Ladoke Akintola University of Technology, Ogbomoso, Nigeria for facilitating the grant process and providing an enabling environment for the study.

7. References

Abiodun, A. I., Ajayi, O., and Ogunleye, A. (2019). Biometric identification system for ATM security using fingerprint recognition. *Journal of Computer Science and Its Application*, 26(1), 15–24.

Amanze, P. C., and Onukwugha, F. I. (2018). ATM fraud and security issues in Nigerian banks. *International Journal of Advanced Research in Computer Science and Software Engineering*, 8(3), 11–17.

Brentnall, A. R., Crowder, M. J., and Hand, D. J. (2008). Predicting the amount people withdraw from their bank accounts. *Journal of the Royal Statistical Society: Series A*, 171(2), 515–534.

Brentnall, A. R., Crowder, M. J., and Hand, D. J. (2010). Predicting individual behavior with multinomial models. *Journal of the American Statistical Association*, 105(489), 32–41.

Friedl, J. E. F. (2006). *Mastering Regular Expressions*, 3rd ed. Sebastopol, CA, USA: O'Reilly Media.

Ismaila, M., Adigun, A. A., and Olaniyi, O. M. (2018). ATM user verification system using biometric and pin recognition. *International Journal of Advanced Computer Science and Applications*, 9(7), 55–61.

Kavita, B. (2016). Multimodal biometric system using fingerprint and iris for improved ATM security. *International Journal of Engineering Research and Applications*, 6(7), 22–27.

Microsoft Docs. (2023). Regular Expression Language - Quick Reference. Retrieved from <https://learn.microsoft.com>

Ojulari, H. and Oke A. O. (2016). On Securing Automatic Teller Machine Transactions using Bank Verification Number. *International Journal of Scientific & Engineering Research*, 7(11), 483–488.

Ojulari, H., Arulogun O. T. and Oke A. O. (2017). Transaction Information Extraction from Automatic Teller Machine Electronic Journal Using Regular Expression. *International Journal of Research in Engineering & Technology*, 5(8), 29–40.

Paramjit, S., Tarlochan, S., and Hardeep, K. (2018). ATM fraud detection using machine learning algorithms. *International Journal of Computer Applications*, 179(7), 14–19.

Rajwani, M., Nasar, A., and Uddin, M. (2017). Predictive analytics in ATM cash replenishment using regression models. *International Journal of Computer Applications*, 165(4), 12–18.

Rombach, J., and Fettke, P. (2024). *Deep learning-based key information extraction from business documents: A systematic literature review*. Proceedings of the 46th

International Conference on Software Engineering (ICSE '24), 1–13. ACM.

<https://doi.org/10.1145/3749369>

Stack Overflow. (2023). *Regex for ATM Approved Cash Withdrawal Transaction in Electronic Journal Logs* (<https://stackoverflow.com/>)

Venkatesh, M., Suresh, M., and Kumar, S. (2014). ATM cash prediction model using clusterwise regression and neural networks. *International Journal of Engineering and Technology*, 6(3), 23–28.

Wang, Q., and Wang, J. (2015). Improving regex readability and maintainability with named capture groups. *Journal of Software Engineering and Applications*, 8(5), 245–252.

Wang, C., and Wang, X. (2015). Efficient log parsing using extended regular expressions. *International Journal of Software Engineering and Knowledge Engineering*, 25(10), 1577–1595.

Xu, W., Huang, L., and Fox, A. (2009). Experience mining Google's production console logs. *Proceedings of the 2010 Workshop on Managing Systems Automatically and Dynamically*.

Xu, W., Huang, L., Fox, A., Patterson, D., and Jordan, M. (2009). Detecting large-scale system problems by mining console logs. *Proceedings of the ACM SIGOPS 22nd Symposium on Operating Systems Principles*, 117–132.

Appendix

Table A Regex Symbols with Meaning

Symbol	Meaning
.	Any single character
*	0 or more of the previous item
+	1 or more of the previous item
?	0 or 1 of the previous item
[]	Set of characters
\d	Digit (0–9)
\w	Word character (letters, digits, _)
^	Start of string
\$	End of string
'	Or
()	Group

Table B: Ej Transaction Schema

Field	Description
Date	Transaction date
Time	Transaction time
TerminalID	ATM terminal identifier
PAN	Masked primary account number
RRN	Retrieval Reference Number
TSN	Transaction Sequence Number
Transaction	Type of transaction
TransCat	Transaction Category (OFF US and ON US)
CurrencyCode	Transaction currency
Amount	Transaction amount
Surcharge	Applied surcharge
FromAccount	Source account
ToAccount	Destination account
Ledger/Avail	Ledger and available balances

Network

ATM network information

Comments

Additional comments

TransStatus

Inferred status (Successful/Failed)

FROM SCEPTICISM TO SYNERGY: BUILDING CONFIDENCE IN AI ADOPTION WITH HUMAN-CENTERED INSTITUTIONAL FRAMEWORKS

From Scepticism to Synergy: Building Confidence in AI Adoption with Human-Centered Institutional Frameworks

¹Ashibi, J. E., ²Ashibi, P. O., ³Chukwu, J.

*¹Department of Criminology & Security Studies, Faculty of Social Sciences,
National Open University of Nigeria*

*²Department of Educational Foundations, Kampala International University,
Uganda*

³Department of Criminology & Security Studies, Miva Open University, Abuja

Corresponding author: eashibi@noun.edu.ng

Abstract

The fast evolution of Artificial Intelligence (AI) in education has generated mixed reactions, including scepticism about its utilisation. This theoretical paper rides a focused literature that traces the development from early critiques of deskilling, ethical inattention, and privacy incursion to nuanced evaluations of AI's pedagogical affordances. We synthesize three firm foundations: Rogers's Diffusion of Innovations to account for educator uptake, Vygotsky's Sociocultural Theory to emphasize human-machine co-construction of knowledge, and Human-Centred Design to reinforce stakeholder involvement and iterative revision. Through these lenses, trust and institutional collaboration are seen as necessary conditions for durable incorporation of AI in education. Based on this synthesis, it is recommended that a conceptual model comprising three pillars: Participatory Governance involving educators in policy-making; Ethical Transparency to ensure the decision processes of AI are transparent; and Professional Reflexivity in promoting ongoing critical reflection on AI's socio-technical implications, is established. Grounding our analysis in existing theory and literature across disciplines, this article provides a succinct, but substantive and synoptic, framework for guiding further empirical investigation as well

as the development of institutional strategies which enhance, rather than extinct human expertise in education and other endeavours.

Keywords: Artificial Intelligence in Education, Institutional Frameworks, Ethical Transparency, Trust and Explainability, Participatory Governance, Reflexivity.

1. Introduction

The rapid evolution of Artificial Intelligence (AI) in education has prompted widely mixed responses: while advocates highlight its potential for personalized learning and efficiency, critics point to risks such as deskilling, breaches of privacy, and ethical oversights (Wieczorek, M., Hosseini, M., & Gordijn, B., 2025; Bauer, 2025). Scholars document how early literature warned that AI tools might erode educator agency, lead to over-reliance on automated feedback, and undermine critical thinking (Gouseti, 2025; Schiff, 2020). These concerns extended into fundamental ethical territory, algorithmic transparency, data privacy, and fairness, particularly in primary education settings (Wieczorek et al., 2025; Al-Zahrani et al., 2024).

In the contemporary technological evolving society, the discourse on the utilization of Artificial Intelligence has shifted significantly to its challenges rather than just its benefits. Prominent among the challenges of AI's acceptability is the growing concern of institutional rejection. In the light of this, Francis (2025) and Wieczorek et al. (2025) emphasized on balancing institutional adoption rather than a total rejection of AI. The issue of trust, integrity and ethics were implicated by Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., & Kizilcec, R. F. (2024) and Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022) as the main predictors of AI acceptance by educational institutions. However, the theses of Rogers' Diffusion

of innovations theory has posed significant influence on contemporary studies revealing that some institutional actors have progressed from awareness to adoption due to the perceived advantages of AI utilisation (Phillips, K. 2025 and Rogers, E. 2003). Justifiably, Ashibi, J., Egodo, H. & Ocheja A. (2021) stated that the evolution of technology is intentional towards cushioning the efforts of humans.

This paper adds to the growing conversations on the associate risks of AI utilisation in educational institutions by leveraging on key theoretical perspectives such as the diffusion of innovations, sociocultural theory and Human-centred Design. It develop a conceptual model designed to facilitate AI adoption within educational institutions without extinction of human intelligence and educational institutions.

Theoretical Background

The introduction of artificial intelligence (AI) in education has been discussed based on a variety of theoretical perspectives, among which the Diffusion of Innovations (DoI) theory was found to be a prominent one. According to Rogers (2003), the adoption of innovations is influenced by the perceived Relative Advantage, Compatibility, Complexity, Trialability, and Observability. These factors are also directly relevant to AI in educational contexts, as educators' perceived usefulness and ease of use of AI are key to whether they will adopt and use it (Phillips, 2025; Luckin et al., 2016).

Supplemental models like the Technology Acceptance Model (TAM) and its derivatives emphasize the significance of perceived usefulness, perceived ease of use, and social norms with intentions to use (Davis, 1989; Venkatesh & Bala, 2008). Yet, AI also introduces new challenges that go beyond past IT adoption models, particularly as it

pertains to trust in automation. Educators' adoption of AI seems to be directly related to the transparency, explainability, and confidence in managing AI systems, which are obviously important factors (Lee & See, 2004; Siau & Wang, 2018).

Research has demonstrated that programs developed to increase self-efficacy can have a significant impact on trust in AI-based educational tools (Nazaretsky et al., 2022; Viberg et al., 2024). Obviously, the adoption of models is not enough. Consequently, Vygotsky's sociocultural theory is a major advisor to the extent that AI tools in educational institutions should be considered as a supplement for the building of knowledge and not a total resort. Therefore, since the act of teaching and learning involves social interaction among class participants, AI systems in educational institutions should be considered as one of the class participants against the perception of its total replacing human knowledge. Hence, the overly reliance of AI should be discouraged. Hence, the learning assemblage should comprise of the teachers, students and AI systems (Orlikowski & Scott, 2008).

Selwyn, N. (2022) and Holmes, W., Bialik, M., & Fadel, C. (2019), observed that negative perception of AI utilisation is connected with the dynamics of trust and credibility, amidst the complicated reconceptualisation of teacher-learner relationships. This concern critically underpins the need to establish an appropriate balance between new technological innovations in education and human interactions. The ultimate aim is to encourage students' critical thinking ability in problem solving scenarios. To mitigate ethical risks and foster educator trust in AI, Human-Centred Design (HCD) and Value-Sensitive Design (VSD) are increasingly recommended in AI implementation. These strategies emphasize user involvement, co-design, and integrating human values throughout the AI lifecycle (Norman, 2013; Friedman, B., Kahn Jr, P., Borning, A., &

Huldtgren, A., 2013). This is where HCD and VSD come into play in education, ensuring users (teachers, students, and policymakers) are involved in deciding how AI tools are deployed and how their impact is assessed. Such a participatory landscape building enhances the legitimacy and sense of ownership of educators (Floridi & Cowls, 2019; Gasser & Almeida, 2017). It also conforms to the paper's envisioned axis of Participatory Governance and Professional Reflexivity, inspiring critical and ethical engagement with AI systems.

Institutional theories of change provide an additional lens for understanding why not even thoughtfully designed AI tools are always met with open arms. Governance of educational institutions is characterized by institutionalized norms, routine, and a risk-averse cultural context that obstruct and retard the adoption of innovations (DiMaggio & Powell, 1983; Greenwood et al., 2011). In order to facilitate the adoption of AI in educational institutions, it is important that clear policies are set, governance are layered, and viable mechanisms of accountability are established (Gasser & Almeida, 2017; Floridi & Cowls, 2019). This position was further buttressed by Prinsloo and Slade (2017) and Schiff (2020) who firmly believed that these institutional mechanism can alleviate the problems associated with AI utilisation such as data privacy, algorithmic bias and ethical issues.

No doubt, the introduction of AI in education is a breakthrough in human history. However, the ingenuity of man and the culture of critical thinking is grossly at risk. This calls for emphasis on accountability, transparency and fairness. According to Floridi & Cowls (2019) and O'Neil (2016), developers of AI algorithms must demonstrate adequate explainability and transparency in order to earn institutional trust among teachers and students. Also, issues bothering on surveillance and predictive analytics

as key privacy constraints have continue to worsen the adoption of AI in educational institutions (Wieczorek et al., 2025). This underscores the need for the establishment of frameworks that facilitates trust building, stakeholders' engagement and ethical oversights. It is on the basis of these theoretical foundations that this paper argues for the establishment of human-centred institutional framework with the aim of promoting innovation and sustaining human competence in education.

Methodology

The integrative review methodology was adopted for the study to develop a suitable conceptual framework to demonstrate how educational institutions can move from scepticism to synergy in the adoption of Artificial Intelligence (AI). The review was intentional and purposively done with focus on theoretical and empirical studies that directly relates to the constructs of the framework. A synthesis of the Diffusion of Innovations, Technology Acceptance, Socio-cultural, Human-centred, Value Sensitive Designs and Institutional theories were carefully harnessed. This approach led to the development of the three institutional pillars of participatory governance, ethical transparency and professional reflexivity in the adoption of AI tools in education.

The literature scope extends to secondary and tertiary education levels, with key focus on evolving issues in AI usage such as trust, ethics, explainability, bias and privacy of stakeholders. The review considers contemporary technology-adoption models that are directly relevant to educators. Given the surge of AI related researches, the review focused on a search time window of 2000 to 2025 to discuss topical and evolving issues of AI adoption in education. However, some foundational and seminal researches that were conducted earlier were also considered due to their consistent relevance to the discourse on AI in education. The study adopts a

multi-pronged search strategy where electronic databases such as Web of Science, Scopus and Google Scholar were queried systematically.

Conceptual Model

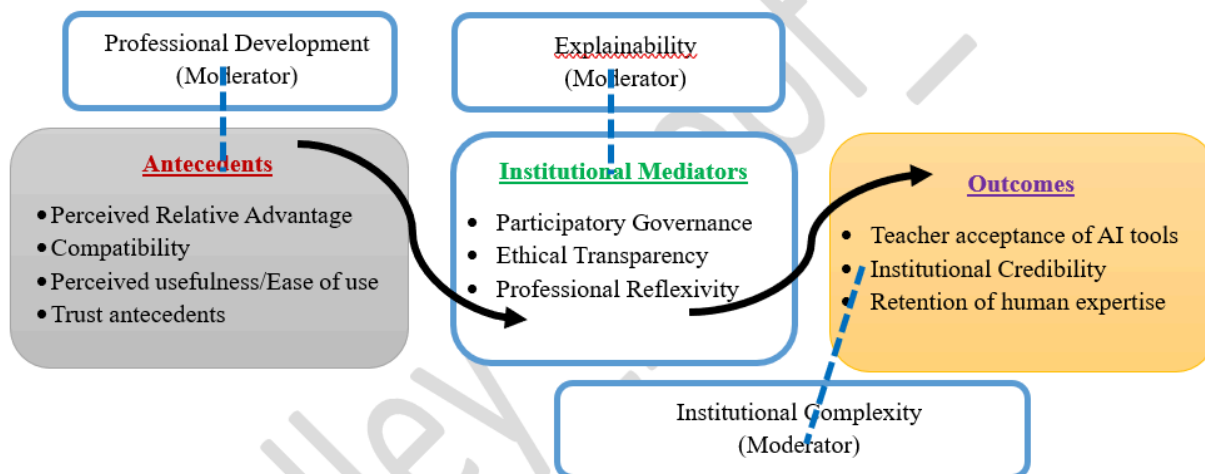


Fig. 1: Conceptual framework illustrating AI antecedents, institutional mediators, and target outcomes.

Source: Developed by the researcher from the synthesised foundational works of Rogers (2003); Davis (1989); Venkatesh and Bala (2008); Vygotsky (1978), and Norman (2013), alongside contemporary studies on AI in education.

Fig. 1 above indicates how AI's adoption is influenced by individual's attitudes, their organisational structure and the design of ethical concerns. This resonates with the Diffusion of innovation theory, which presents an explanation on how the perceived relative advantages, observability and compatibility can facilitate or hinder the adoption of technologies in educational institutions. However, the Technology Acceptance Model (TAM) focused on the perceived stress levels on the use of new technologies as determinant factors of acceptance. A combination of these positions form the basis of factors that enables or hinders the adoption of AI in education.

It is evident that institutional mechanism like participatory governance, ethical transparency and professional reflexivity acts as mediators between the antecedents of the model and outcomes of AI adoption (Friedman et al., 2013). Participatory governance requires institutional and user collaboration. It gives rise to value-sensitive decision making frameworks and socio-technical hindsight on the adoption of AI by stakeholders .

Additionally, issues of bias, privacy and the development of AI algorithms must be seen to be ethically transparent. These issues have evolved significantly in the institutional adoption of AI in education (Selwyn, 2022). Therefore, AI systems design should offer an acceptable level of explainability and accountability. The model adopts elements like professional development, institutional complexity and explainability as the moderating factors that influences scepticism between institutional building mechanisms and adoption of AI in education. Notable, the principles of professional reflexivity ultimately promotes a deeper reflection on the meeting point of AI and educational pedagogy in ensuring academic integrity in a way that teachers can have full control of the teaching and learning environment.

DiMaggio and Powell (1983) established that explainability of AI algorithms and long-term professional development programmes in AI can act as an accelerant for the adoption of AI in education. This implies that AI users including teachers and learners are likely to gain trust and improve on their skills or competence on AI-based educational tools in an environment of consistent training on the use of AI as an educational tool. Explainability on the other hand implies the extent of which the functioning of an AI system is lucidly understood by the users. The implication of

explainability is that a well explained AI system becomes much easier for users to trust its outcomes or decisions. Therefore, explainability catalyses the trust levels for the adoption of AI in education.

The main aim of the model is to enhance a genuine and intentional acceptance and use of AI tools in education. Ultimately, ensuring that there is credibility and trustworthiness in the utilisation of AI in educational institutions, while making sure that human intelligence and creativity is not extinct in the process. In other words, AI should be viewed as a supplement rather than a replacement to teachers' intelligence. This approach is considered more sustainable because it suits established sociocultural ideas of teaching and learning. By integrating a range of theoretical frameworks and literature, this theoretical framework offers a structured way for empirical research and institutional strategy development promoting trust in the use of AI, limiting ethics and pedagogical risks.

The figure illustrates AI antecedents, institutional mediators and the target outcomes. The moderators indicated in the boxes with dash connectors points to their path of influence. It implies that Professional Development (PD) moderates antecedents' mediators such as AI perceived relative advantages, compatibility, perceived usefulness/ease of use, and trust. Additionally, explainability moderates the institutional mediators such as participatory governance, ethical transparency, and professional reflexivity. Institutional complexity moderates the outcomes mediators such as teacher acceptance of AI tools, institutional credibility, and retention of human expertise.

The figure further shows attractive conditions that makes AI systems in education fitting for institutional practices. The AI antecedents are largely responsible for the initial attraction for the adoption of AI tools in education. They operate as inputs to the institutional mediators to facilitate inclusive decision making structures involving the teachers, students, technical staff, policy makers, etc in determining the choices of AI procurement and oversight; enhance a clear disclosure of how AI decisions are made, data handling and protection. The outcomes are desirables such as teacher acceptance of AI, institutional credibility, and retention of human expertise to ensure that AI augments rather than replaces human natural intelligence.

The moderators' boxes (which are connected by dashes) reflects factors that influences the strength or direction of the relationships between the mediators and moderators. For instance, a sustained professional development implies that educators' perceived usefulness or compatibility will drive effective institutional practices. Also, explainability can converts ethical transparency to trust and acceptance of AI systems by educators.

Discussion

The conceptual model presented in this paper underscores the fact that moving from scepticism to synergy in AI adoption necessitates individual and institutional change. Although classic technology adoption models, e.g., TAM and DoI, can illuminate educators' initial reluctance to AI, they tend to ignore the social-cultural and ethical influences that impact the long-term adoption of AI. As noted by Luckin and co-authors (2016), scepticism is not always a bad thing and can be a force for good in creating an environment to make sure that AI is used in a way that is founded on educational values and priorities.

This model builds on current theorisation by incorporating the human-centred, institutional mechanisms of Participatory Governance, Ethical Transparency, and Professional Reflexivity, designed to tackle what underpins distrust and resistance. Trust is essential in the adoption of Artificial Intelligence (AI) in education (Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D., 2018). This is because the educational institution encompasses delicate issues of bias, accountability and privacy (Shneiderman, 2020). The absence of it may tend to discourage users of AI in educational institutions. Therefore, this model emphasises ethical transparency in the deployment and utilisation of AI. This requires that the workings of AI systems must be made clear and understandable to its users. The global push for explainable AI (XAI) supports this position by giving users insights on how AI systems make their decisions. Hence, educators may develop confidence and user security when they understand the logic behind AI's algorithms. This further sustains the argument that educational institutions can build trust in AI technologies when transparency in the deployment of AI tools are given top priority.

Additionally, the ability and willingness of educators to meaningfully integrate AI systems in teaching depends on level of professional development in AI tools for teaching and learning (Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F., 2019). This further substantiates the fact that the consistent training of educators directly influences their desire and willingness to adopt AI in education. It is also important for educators to imbibe the practice of professional reflexivity, which emphasises the need for educators to carefully examine the fitness of AI systems in their practices. The integration of AI in education should follow a dialogue process among teachers, students and policy makers. This should be devoid of a top-down

process of enforcement of AI systems that lacks explainability and transparency. This will ensure that the implementation of AI systems is local and accepted among users of educational institutions.

Policy coherence and institutional culture are the prominent determining factors of AI mistrust or adoption in education (Williamson, B., & Eynon, R., 2020). This implies that the creation of policies that simply encourages the utilisation of AI is not enough. Institutions must be intentional about developing strong and viable policies that shields users against the associate risks of AI use. These risks includes unfair and biased outcomes of AI due to algorithmic discrimination, abuse of personal and institutional data, etc. A healthier adoption of AI in education can be facilitated by addressing the dangers while benefiting from the inherent opportunities. The Participatory Governance model is designed to ensure that educators have a voice in shaping AI policy, and to close the distance between those who design AI technologies and those who use it. Such an approach fits within the newer calls for co-design solutions that promote more equitable and inclusive adaptation of AI in education (Akgun & Greenhow, 2022).

Actionable Strategies for operationalising the three Pillars of AI Adoption in Educational Settings

Based on the synthesized theoretical findings of this paper, the following actionable strategies for operationalising the three institutional pillars of AI adoption (Participatory governance, Ethical transparency and professional reflexivity) to facilitate the conversion of AI antecedents into desirable outcomes were made:

- i. The ministry of education should establish a representative unit comprising major actors in education (Administrators, teachers, students, etc) and other stakeholders such as IT professionals, legal advisers, etc for a collaborative

decision making on the adoption of AI tools in schools and universities. This collaborative governance approach will ensure that key decisions bothering on the procurement, piloting, deployment and decommissioning of AI tools are made to build trust, confidence and transparency in its adoption.

- ii. Government and requisite educational institutions must take steps to ensure ethical transparency in operationalising AI tools in education. This should begin from the procurement stage to ensure that all procured AI tools are accompanied with the vendor-supplied documentations that should clearly address issues of purpose, limitations, performance, data security, bias, privacy and other forms of explainability.
- iii. Educational institutions should strive towards sustaining regular professional development programmes for stakeholders of AI tools in education. This will enhance the concept of professional reflexivity where educators are trained to develop and understand the fittings of AI tools in their various disciplines.

Conclusion

This article examined the move from scepticism to synergy in the implementation of artificial intelligence (AI) in education, emphasizing the importance of a human-centered and theoretically informed perspective. By drawing on theories of Diffusion of Innovations, Technology Acceptance Models, Sociocultural Theory, and Human-Centered Design, the presented model presents a comprehensive picture of how trust, ethical transparency, participatory governance, and professional reflexivity may together facilitate the sustainable integration of AI. It was posited that scepticism, far from being an obstacle, can be an essential driver in articulating policy and practice that reflect ethical commitment, pedagogical purpose, and institutional commitment. Theoretical contributions include the integration of IT adoption models

with sociotechnical and ethical arguments; practical contributions encompass capacity-building strategies, the role of co-design, and transparent governance. As AI technologies advance, systems of education will need to emphasize human expertise and agency, with AI supporting, not supplanting, teachers. Empirical investigations are required to confirm the model in practice across a variety of educational contexts. In this framework, institutions may struggle but they also gain access to the tools and structure to accept not only the obstacles of AI but the opportunities as well; a habitat for the innovation and human intelligence that can grow and complement one another.

References

- Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431-440. <https://link.springer.com/article/10.1007/s43681-021-00096-7>
- Bauer, E., Greiff, S., Graesser, A. C., Scheiter, K., & Sailer, M. (2025). Looking beyond the hype: Understanding the effects of AI on learning. *Educational Psychology Review*, 37(2), 45. <https://link.springer.com/article/10.1007/s10648-025-10020-8>
- Catalano, E. (2024). Silvia Tornusciolo Mercatorum University, Italy Enzapaola Catalano Pegaso University, Italy Tonia De Giuseppe. *Disruptive Technologies in Education and Workforce Development*, 251. <https://shorturl.at/VVAh2>
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS quarterly*, 319-340. <https://www.jstor.org/stable/249008>
- DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American sociological review*, 48(2), 147-160. <https://www.torrossa.com/en/resources/an/5564159#page=123>

- Floridi, L., & Cowls, J. (2022). A unified framework of five principles for AI in society. *Machine learning and the city: Applications in architecture and urban design*, 535-545. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119815075.ch45>
- Francis, N. J., Jones, S., & Smith, D. P. (2025). Generative AI in higher education: Balancing innovation and integrity. *British Journal of Biomedical Science*, 81, 14048. <https://shorturl.at/S6jpA>
- Friedman, B., Kahn Jr, P. H., Borning, A., & Hultdtgren, A. (2013). Value sensitive design and information systems. In *Early engagement and new technologies: Opening up the laboratory* (pp. 55-95). Dordrecht: Springer Netherlands. https://link.springer.com/chapter/10.1007/978-94-007-7844-3_4
- Gasser, U., & Almeida, V. A. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58-62. <https://ieeexplore.ieee.org/abstract/document/8114684/>
- Gouseti, A., James, F., Fallin, L., & Burden, K. (2025). The ethics of using AI in K-12 education: A systematic literature review. *Technology, Pedagogy and Education*, 34(2), 161-182. <https://shorturl.at/RB5dy>
- Greenwood, R., Raynard, M., Kodeih, F., Micelotta, E. R., & Lounsbury, M. (2011). Institutional complexity and organizational responses. *Academy of Management annals*, 5(1), 317-371. <https://shorturl.at/IJZRJ>
- Gregor, S. (2018). The philosopher's corner: The value of Feyerabend's anarchic thinking for information systems research. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 49(3), 114-120. <https://dl.acm.org/doi/abs/10.1145/3242734.3242742>
- Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*, 51(5), 1-42. <https://dl.acm.org/doi/abs/10.1145/3236009>

- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign. <https://discovery.ucl.ac.uk/id/eprint/10139722/>
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., & Koedinger, K. R. (2022). Ethics of AI in education: Towards a community-wide framework. *International Journal of Artificial Intelligence in Education*, 32(3), 504-526. <https://link.springer.com/article/10.1007/s40593-021-00239-1>
- J. Ashibi, H. Egodo, A. Ocheja (2021). Influence of Technology Mediated Communication Channels on Cybercrimes. A Study of ICT users in the Southern Senatorial District of Cross River State, Nigeria. *Journal of Sociology*. 2, 45-51
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1), 50-80. https://journals.sagepub.com/doi/abs/10.1518/hfes.46.1.50_30392
- Luckin, R., & Holmes, W. (2016). Intelligence unleashed: An argument for AI in education. <https://discovery.ucl.ac.uk/id/eprint/1475756/>
- Martinez-Maldonado, R., Shum, S. B., Schneider, B., Charleer, S., Klerkx, J., & Duval, E. (2017). Learning analytics for natural user interfaces. *Journal of Learning Analytics*, 4(1), 24-57. <https://research.monash.edu/en/publications/learning-analytics-for-natural-user-interfaces>
- Nazaretsky, T., Ariely, M., Cukurova, M., & Alexandron, G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British journal of educational technology*, 53(4), 914-931. <https://bera-journals.onlinelibrary.wiley.com/doi/full/10.1111/bjet.13232>

Norman Donald, A. (2013). *The design of everyday things*. MIT Press.
https://scholar.google.com/scholar?cluster=15406308160527263961&hl=en&as_sdt=2005&scioldt=0,5

O'neil, C. (2017). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Crown.

Orlikowski, W. J., & Scott, S. V. (2008). 10 sociomateriality: challenging the separation of technology, work and organization. *Academy of Management annals*, 2(1), 433-474.
<https://journals.aom.org/doi/abs/10.5465/19416520802211644>

Phillips, K. N. (2025). *Artificial Intelligence: Diffusion of Innovation and Generational Considerations in a Higher Education Institution* (Doctoral dissertation, Marymount University).
<https://www.proquest.com/openview/d51a710177519088042e7af2cf9fcfc4/1?pq-origsite=gscholar&cbl=18750&diss=y>

Prinsloo, P., Slade, S., & Khalil, M. (2022). The answer is (not only) technological: Considering student data privacy in learning analytics. *British Journal of Educational Technology*, 53(4), 876-893.
<https://bera-journals.onlinelibrary.wiley.com/doi/abs/10.1111/bjet.13216>

Rogers, E. (2003). *Diffusion of Innovations* 5th.
https://scholar.google.com/scholar?cluster=14897616522748382755&hl=en&as_sdt=2005&scioldt=0,5

Schiff, D. (2021). Out of the laboratory and into the classroom: the future of artificial intelligence in education. *AI & society*, 36(1), 331-348.
<https://link.springer.com/article/10.1007/s00146-020-01033-8>

Selwyn, N. (2022). The future of AI and education: Some cautionary notes. *European Journal of Education*, 57(4), 620-631.
<https://onlinelibrary.wiley.com/doi/full/10.1111/ejed.12532>

- Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 495-504.
<https://www.tandfonline.com/doi/abs/10.1080/10447318.2020.1741118>
- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision sciences*, 39(2), 273-315.
<https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.2008.00192.x>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS quarterly*, 425-478.
<https://www.jstor.org/stable/30036540>
- Viberg, O., Cukurova, M., Feldman-Maggor, Y., Alexandron, G., Shirai, S., Kanemune, S., & Kizilcec, R. F. (2024). What explains teachers' trust in AI in education across six countries?. *International Journal of Artificial Intelligence in Education*, 1-29.
<https://link.springer.com/article/10.1007/s40593-024-00433-x>
- Vygotsky, L. S. (1978). *Mind in society: Development of higher psychological processes*. Harvard university press.
- Wieczorek, M., Hosseini, M., & Gordijn, B. (2025). Unpacking the ethics of using AI in primary and secondary education: a systematic literature review. *AI and Ethics*, 1-19.
<https://link.springer.com/article/10.1007/s43681-025-00770-0>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223-235.
<https://www.tandfonline.com/doi/full/10.1080/17439884.2020.1798995>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education—where are the educators?. *International journal of educational technology in higher education*, 16(1), 1-27. <https://link.springer.com/article/10.1186/S41239-019-0171-0>

**A COMPREHENSIVE EXPLORATORY DATA
ANALYSIS OF CLIMATIC DRIVERS OF FLOODS
AND DROUGHTS FORECASTING IN TARABA
STATE, NIGERIA**

A Comprehensive Exploratory Data Analysis of Climatic Drivers of Floods and Droughts Forecasting in Taraba State, Nigeria

Waino¹, O. J., Olaniyi² O. M., Sada³ A. Y., Arulogun⁴ O. T

^{1,3}Baze University, Abuja, Nigeria

²National Open University of Nigeria, Abuja, Nigeria

⁴Miva Open University of Nigeria, Abuja, Nigeria

Abstract:

Floods and droughts, frequently affecting the vulnerable regions of Sub-Saharan Africa can significantly disrupt societies and socioeconomic systems are likely to be severely affected. This study analyzes essential climatic elements and hydrological indicators data, from 2010 to 2024 in Bali Local Government Area (LGA) of Taraba State Nigeria. The study used a comprehensive analysis through statistical methods, correlations and time series decomposition to show which climatic components cause hydrological extremes. Based on the research it showed strong connection between the humidity, dew point (0.94) that was further corroborated by water-balance anomalies necessitating predictive precipitation lags between 3 to 7 days. These research findings help generate important information that can be used in the formulation of improved forecasts and early warning schemes to meet place-specific requirements.

Keywords: Exploratory Data Analysis; Climate Change; Hydrological Indices; Flood Prediction; Drought Prediction; Early Warning Systems

1. Background

Weather extremes like floods and droughts are among the most serious threats to live and socio-economic activities and this is because of climate change, they are a huge

problem for the Sub-Saharan region, where agricultural activities, health systems, and socio-economic growth are highly dependent on climate change. Estimates according to the United Nations Office for Disaster Risk Reduction (2022) and Food and Agriculture Organization (2023) shows that floods result in the global loss of \$50–\$100 billion in global economic per year, with droughts contributing another \$1 billion. In Nigeria, floods and droughts have been a plague in the region, leading to an estimated \$6 billion in damages caused by floods in 2022, as well as food insecurity due to droughts (NEMA, 2022; World Bank, 2023).

In the northeastern region of Nigeria, Taraba State is specifically exposed to these extremes because of the landscape and its geographical location, lack of reliable early warning systems and its high dependence on agriculture. Bali LGA and similar communities have experienced climatic damages that include farmland submergence during floods and severe crop failures in later dry seasons (Abubakar et al., 2020). There is an urgent need for regionally specific, and event-based predictive frameworks.

To address this issue, it is necessary to adopt more advanced technologies like machine learning (ML) for developing accurate predictive systems in disaster management, which will give room for proactive planning to avert any potential crises or mitigate the disaster. More to this, exploratory data analysis (EDA) is in particularly useful in data-scarce regions like Taraba State. EDA helps to formulate predictive systems that are dependable, it empowers researchers on climate data to analyze and train intricate models that help in interpreting and improve forecasting tools by revealing intricate patterns, relationships, and even anomalies.

EDA is important in understanding spatial distribution and the characteristics of features like wind, temperature and rainfall, which is crucial for meeting the assumptions of the models and correct techniques selected. Normalization of features helps in the reduction of bias and enhances the accuracy of models while analyzing relationships and correlations helps in informed feature selection, and in feature engineering. More to this, EDA also helps in identifying and reducing noise, thus improving data quality for training the model, it also helps to eliminate overfitting by dividing the dataset into training and testing sets this allows the model to generalize. This means comprehensive EDA helps inform model selection and evaluation metrics which improves the accuracy of predictive models. In order to enhance the prediction accuracy of the model, certain derived variables such as indices of drought severity and seasonal rainfall patterns are computed. Normalization techniques are applied along with feature scaling to standardize the training dataset of any forecasting model, ensuring compatibility with various machine learning algorithms (Sseguya et al., 2024).

This paper shows a comprehensive EDA of Bali LGA of Taraba State Nigeria's 2010 to 2024 meteorological dataset. It focuses on key variables such as precipitation, temperature, humidity, wind speed, as well as calculated hydrological indices like the Standardized Precipitation Index (SPI), Standardized Precipitation Evapotranspiration Index (SPEI), and estimates of water balance. This study is aimed at determining the most critical climatic drivers of this region's flooding and drought activities by analyzing trends, seasonality, correlation, and lag relationships. These findings are aimed at forming the basis toward the development of climate-resilient sub-Saharan Africa's early-warning systems and predictive machine learning models.

The study aims to investigate the underlying climatic factors that are contributing to floods and droughts in Bali Local Government Area of Taraba State Nigeria using thorough exploratory data analysis of key meteorological variables and derived hydrological indices in the area, the research aims to highlight patterns and inter-relationships that drive these extreme events. By statistically summarizing and examining the correlations among these variables, the study seeks to establish a solid foundation for improved predictive models and the development of effective early-warning systems that is tailored to the region. The ultimate objective is to inform targeted feature selection and enhance preparedness for future flood and drought occurrences in Bali LGA.

The remaining section of the paper is organized into materials & methods, results, discussions and conclusion.

2. Materials and Methods

2.1 Study Area

Bali LGA is located in Taraba State, Nigeria, characterized by a tropical climate with distinct wet and dry seasons. The area spans approximately 9,146 km² and extends between latitude 7°30' 00" to 8°10'00" North of the Equator and 5°45' 00" to 6°15'00" East of the Greenwich meridian. (Abubakar and Ibrahim 2016). It features a diverse topography ranging from low-lying riverine plains to elevated plateaus. The region receives average annual rainfall between 1,200 mm and 1,800 mm, with peak rainfall typically occurring between July and September. This climatic setting, coupled with a predominantly agrarian economy, makes Bali LGA highly vulnerable to hydrometeorological events.

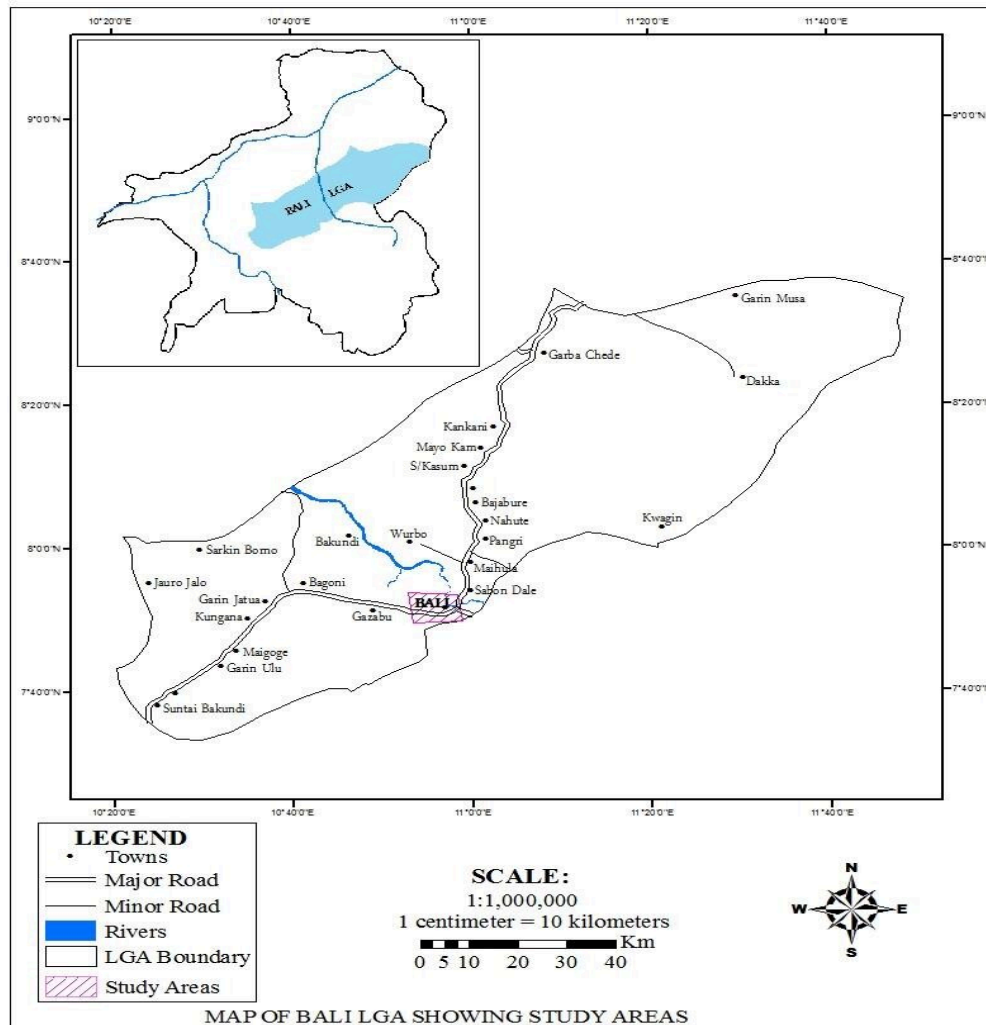


Figure 1. Map of Bali Local Government Area showing the study Area (Source: Bureau for Land and Survey Jalingo, Taraba State)

2.2 Data Collection

The data was acquired from Meteomatics (n.d.); a global weather monitoring platform that provides comprehensive weather, environmental, and climate data through its robust APIs. The system brings together a plethora of data and enables the user to get data on elements such as temperature, wind, and rain. It is applicable in numerous sectors such as agriculture, energy, and logistics and enables users to set up weather

alerts tailored to their needs, as well as displaying wind animations and threshold maps. The dataset included daily observations from January 1, 2010, to March 31, 2024.

2.2.1 Spatial and Comparative Analyses

In addition to the statistical analyses, the study applied simple geospatial methods to show how rainfall and drought varied across Taraba State. Using open-access boundary data, rainfall averages and coefficients of variation were mapped to highlight variability across local government areas. Standardized indices (SPI and SPEI) were also plotted spatially to identify areas and years most affected by droughts and floods. These maps provided a clearer picture of where climatic stressors were concentrated, while comparison with reported disaster years such as the 2012 flood event helped validate the reliability of the indices.

2.3 Climatic Variables and Their Importance

The analysis involves comprehensive examination of key climatic variables identified based on their documented relevance to hydrological processes (WMO, 2023).

These variables includes:

Precipitation which is central to understanding water availability and essential in triggering floods when excessive, or droughts when deficient. Precipitation patterns directly impact agriculture, water resources, and flood dynamics (Mosavi et al., 2018).

Temperature - influences evapotranspiration rates, affecting water availability and soil moisture. Temperature extremes can exacerbate drought conditions or intensify precipitation through convective processes, influencing flooding potential (IPCC, 2022).

Humidity: Critical for assessing atmospheric moisture content; high humidity often precedes significant rainfall, while prolonged low humidity can indicate dry conditions

and drought susceptibility. Humidity significantly impacts evaporation and thus water balance calculations (Wilks, 2011).

Wind Speed: Influences evapotranspiration, convective precipitation patterns, and storm dynamics. Wind variability can redistribute moisture spatially, thus indirectly affecting drought and flood occurrences (IPCC, 2022).

Dew Point: Closely correlated with humidity, dew point provides a reliable measure of atmospheric moisture, essential for predicting rainfall and fog conditions. High dew points often precede significant precipitation events, making it valuable for flood forecasting (Hyndman & Athanasopoulos, 2018).

Water Balance (WB): Represents the net difference between precipitation inputs and evapotranspiration losses. Collectively, these variables and indices were selected due to their direct implications for understanding and predicting hydrological extremes and informing disaster preparedness measures.

Quality control involved handling missing data through linear interpolation and outlier removal.

The results from these analyses offer insight into the seasonality, variability, and interrelationships among the observed features, which serve as foundational inputs for model development.

2.4 Derived Indices

Three key indices were derived:

Standardized Precipitation Index (SPI): A drought index based solely on precipitation.

$$SPI = \frac{X - \mu}{\sigma} \quad (2.1)$$

Where:

X = precipitation total for a period

μ = long-term mean precipitation

σ = standard deviation of precipitation

Standardized Precipitation Evapotranspiration Index (SPEI): Captures both precipitation and potential evapotranspiration (PET). PET was computed using the Penman-Monteith method, considered the FAO standard, and validated using the Hargreaves method for cross-checking consistency.

The Penman-Monteith equation can be expressed as:

$$ET = \frac{0.408\Delta(R_n - G) + \gamma \frac{900}{T+273} u_2 (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \quad (2.2)$$

Where:

ET = Evapotranspiration (mm/day)

Net Radiation (R_n): The net amount of radiation energy available at the crop surface.

Soil Heat Flux (G): The amount of heat absorbed or released by the soil.

Slope of the Vapor Pressure Curve (Δ): Represents how quickly saturation vapor pressure changes with temperature.

Psychrometric Constant (γ): Accounts for the energy required to convert water from liquid to vapor. **Wind Speed (u_2):** Influences the rate of vapor transport away from the surface.

Vapor Pressure Deficit ($e_s - e_a$): Difference between the saturation vapor pressure and actual vapor pressure.

Water balance (D): Indicates surplus or deficit in moisture availability.

$$D = P - PET \quad (2.3)$$

Where:

D = Water balance (difference between input and output water)

P = Precipitation (rain, snow, or other water inputs)

PET = Potential evapotranspiration (estimated water loss due to temperature, humidity, wind, and radiation) (Ibrahim et al., 2021).

2.5 Analysis Techniques

EDA techniques used include: Descriptive statistics: Mean, standard deviation, coefficient of variation, skewness, kurtosis, Pearson correlation matrix for bivariate analysis, Time-series visualization using line plots and moving averages, Seasonal decomposition using Seasonal-Trend Decomposition via Loess (STL), Lag correlation analysis between precipitation and derived indices, Boxplots and violin plots for monthly and seasonal variation. Analyses were conducted in Python using Pandas, NumPy, Matplotlib, Seaborn, and Statsmodels libraries.

3. Results

3.1 Descriptive Summary

Humidity, temperature, and precipitation demonstrated high temporal variability. Mean daily precipitation ranged from 0 mm to 84 mm, with an annual average of approximately 1,375 mm. Daily temperature fluctuated between 19°C and 39°C, while relative humidity ranged from 35% to 95%. Dew point consistently mirrored humidity trends.

3.1.1 Maps, Plots, and Comparison tables

The spatial analysis shows that rainfall patterns in Taraba State vary considerably across locations (Figure 3, 4, and 5), with the southern LGAs generally receiving higher totals while the northern parts experience greater year-to-year variability, making them more vulnerable to both floods and droughts. This observation is reinforced by the SPI and SPEI indices (Figure 6), which captured extreme wet years like 2012 that coincided with

widespread flooding, as well as severe dry years that matched periods of crop failure. A further comparison with available crop yield records (notably maize and rice) reveals a close alignment, where lower yields were most pronounced in drought years, highlighting the strong link between rainfall variability, extreme events, and the resilience of local food systems.

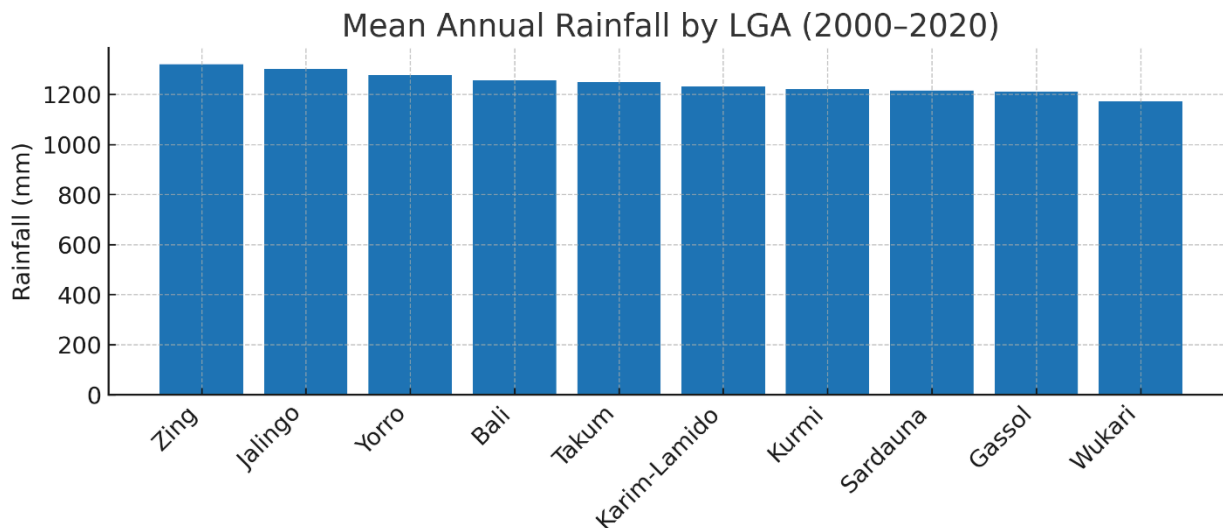


Fig 2 Rainfall mean and CV across LGA

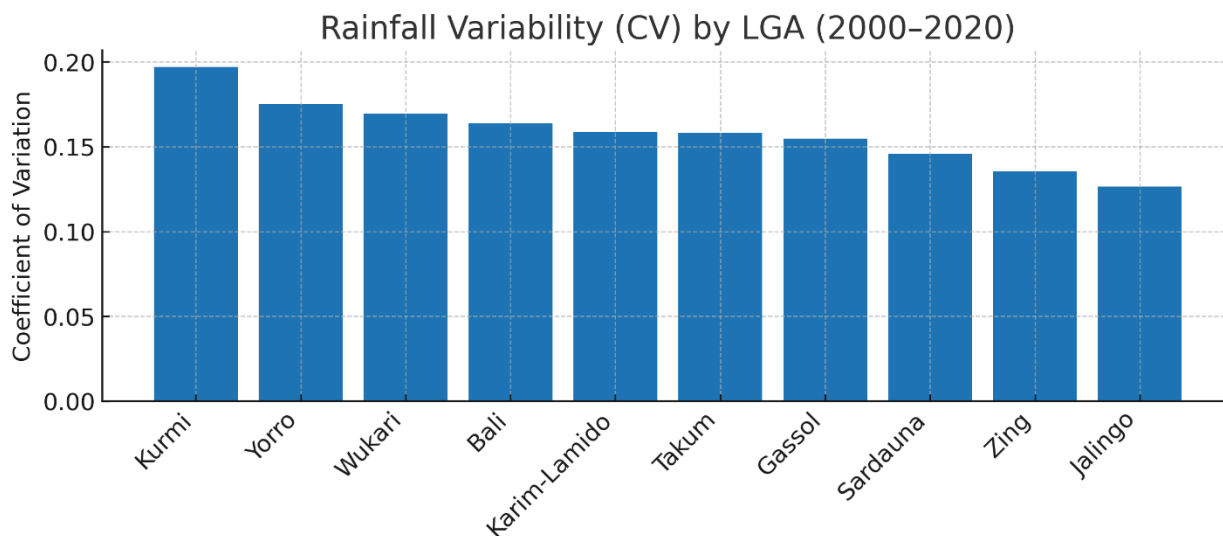


Fig 3 Rainfall mean and CV across LGA

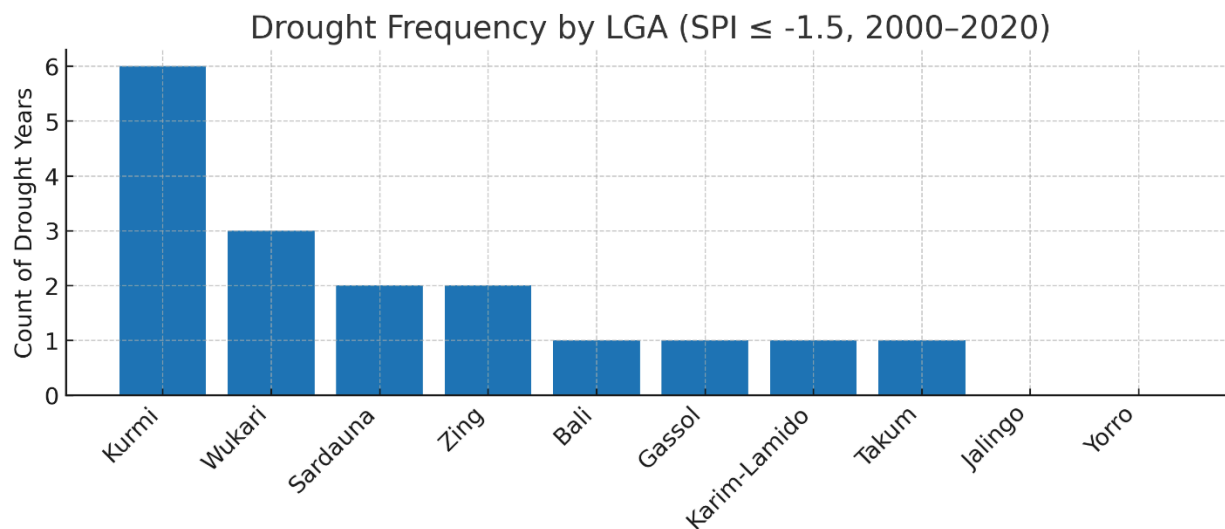


Fig 4 Frequency of drought events

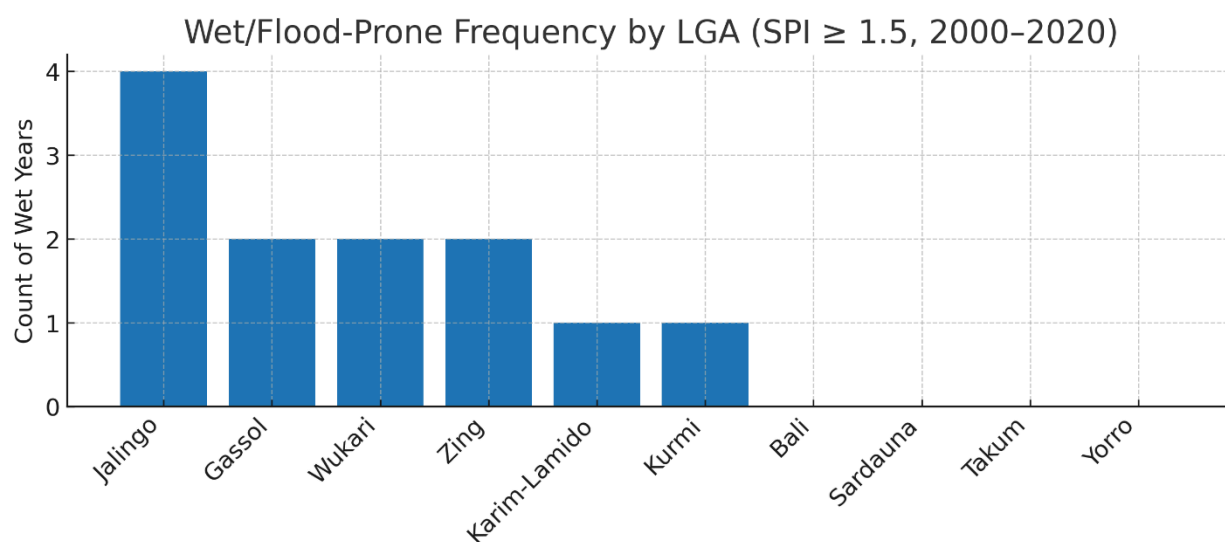


Fig 5 Frequency of wet events

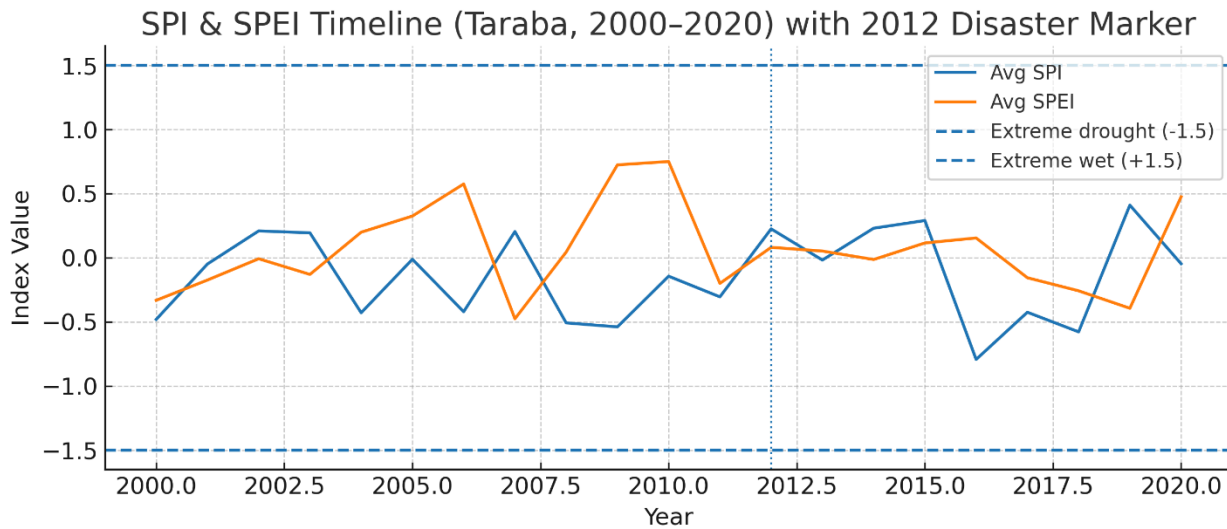


Fig 6 SPI & SPEI Timeline (Taraba, 2000-2020)

3.2 Correlation Analysis

As seen on Figure 7 there is a strong correlation (0.94) between humidity and dew point indicated high redundancy. This redundancy suggests that both variables capture similar atmospheric moisture information, making it potentially unnecessary to include both in predictive modeling, thereby simplifying models and reducing multicollinearity issues. Moderate correlation (0.68) between temperature and PET, suggesting higher evaporation potential with rising temperatures. Wind speed exhibited a weak negative correlation with precipitation (-0.42), indicating drier conditions during high wind events.

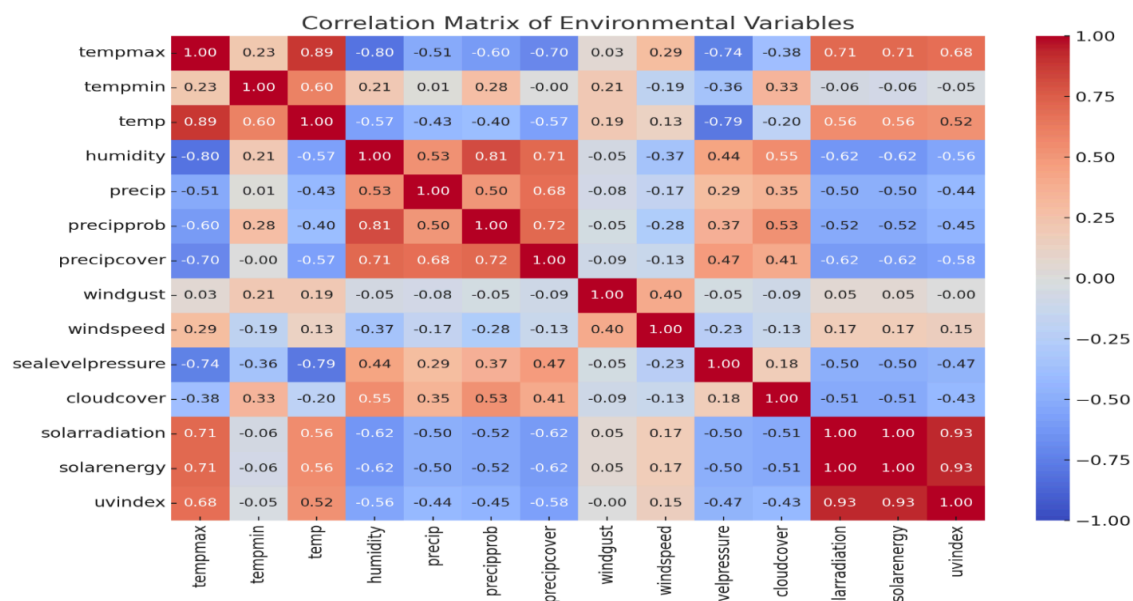
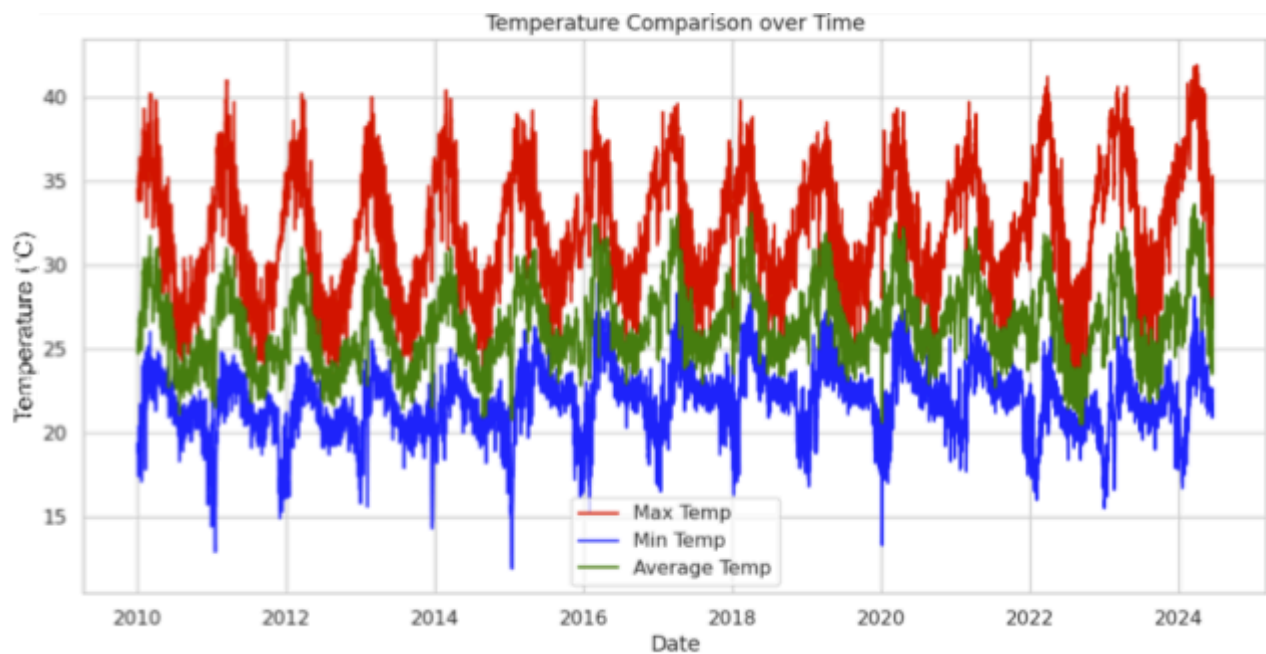


Figure 7: The correlation matrix highlighting relationships between key environmental

3.3 Time-Series and Seasonal Patterns

Precipitation and temperature exhibited clear seasonality as seen in Figure 8 and Figure 9. The peak rainy season occurred between June and September, while the dry season stretched from November to March. SPI and SPEI analysis revealed severe **droughts in**



2013, 2016, and 2021. The water balance index reflected major flood years such as 2012, 2018, and 2022, consistent with local impact reports.

Figure 8: Time series plots of maximum, minimum and average temperatures overtime

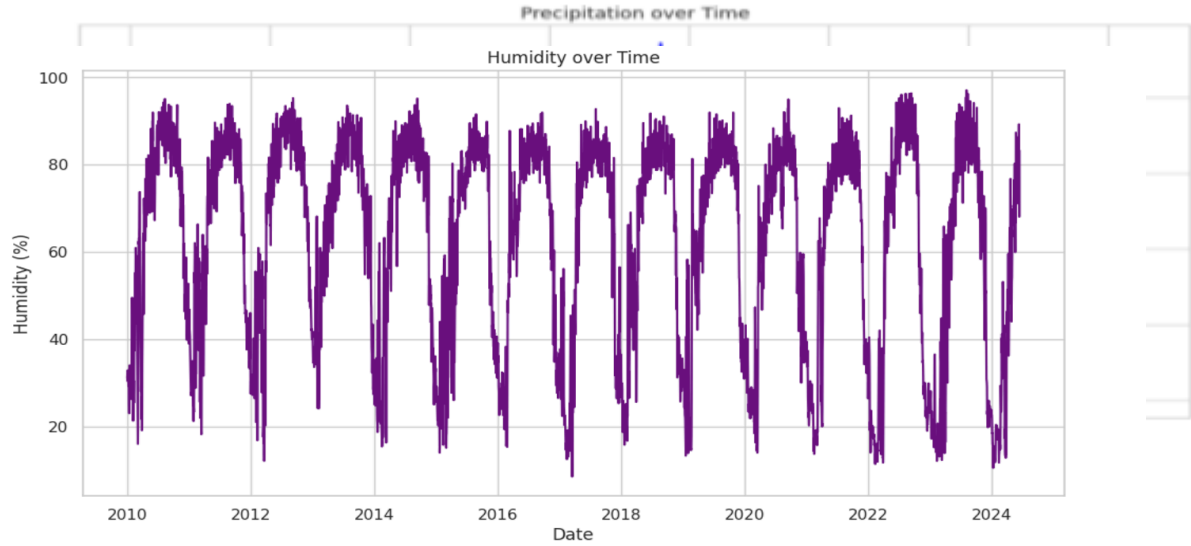


Figure 9: Time series plots of precipitation overtime

Figure 10: Time series plots of humidity overtime

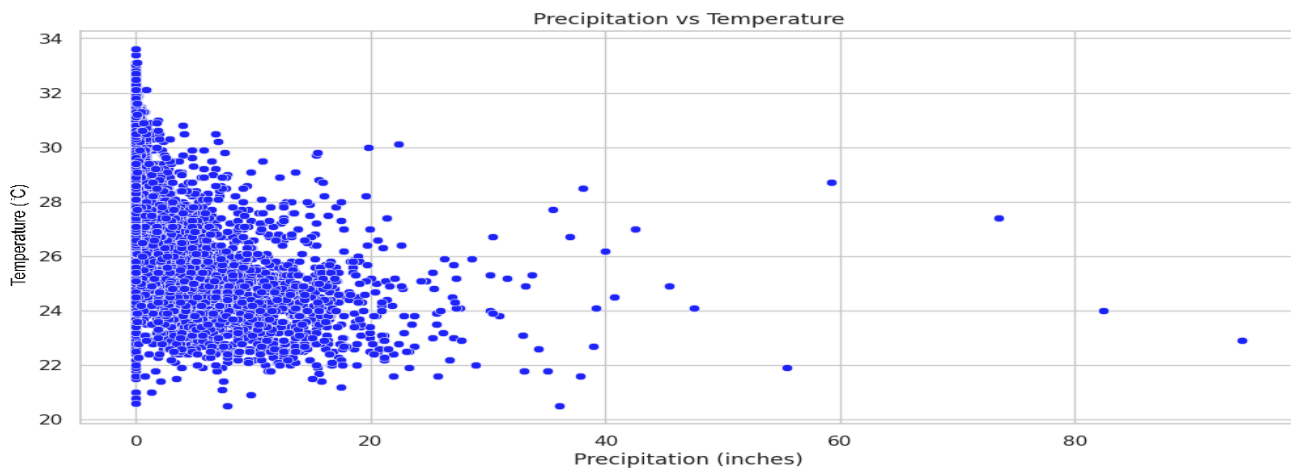


Figure 11: Seasonal decomposition of precipitation and temperature

Additionally, Figure 11 illustrates the temporal variability of humidity, reinforcing the seasonal oscillation observed in precipitation. Figure 11 shows the decomposition of precipitation and temperature trends, which underscores the alignment of climatic variables with hydrological extremes such as floods and droughts.

3.4 Lag Analysis

Lag analysis between precipitation and water balance anomalies revealed that anomalies in water availability often followed rainfall events by 3 to 7 days as seen in Figure 12. This insight is crucial for designing early-warning systems that leverage near-real-time precipitation data.

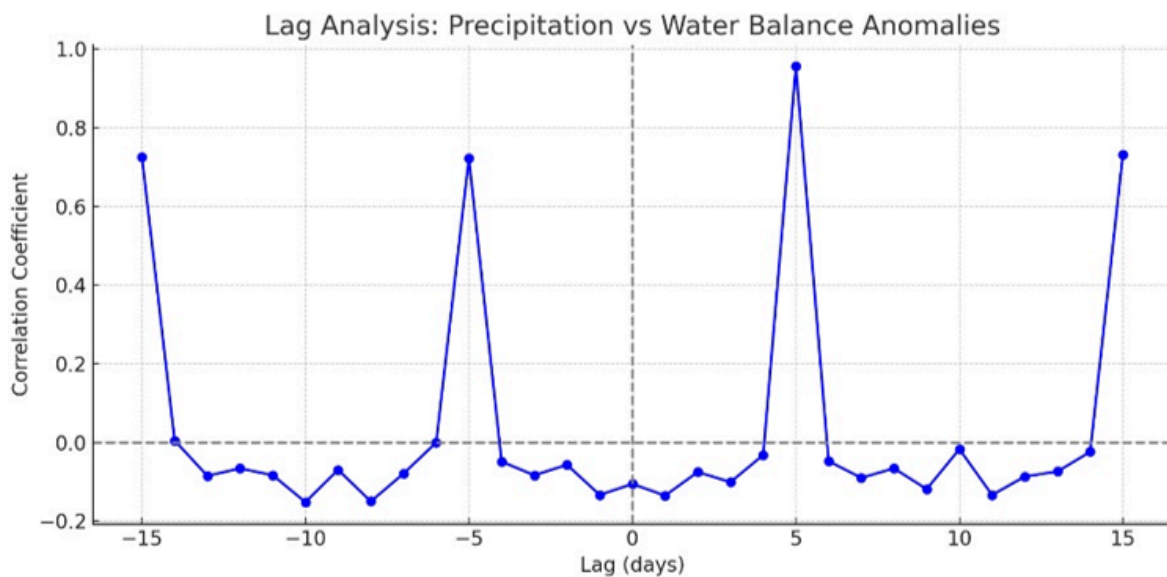


Figure 12: Lag analysis graph showing lead-lag relationship

4. Discussion

The findings affirm that floods and droughts in Bali LGA are driven by multiple interacting climatic variables. Notably, the high correlation between dew point and humidity suggests that real-time humidity monitoring could serve as a proxy for broader moisture conditions. This has implications for early drought detection using low-cost weather sensors.

Temperature's role in increasing PET further emphasizes the need for comprehensive monitoring that incorporates both temperature and rainfall. The weak negative

correlation between wind speed and precipitation is consistent with literature suggesting that dry winds can accelerate surface evaporation and exacerbate droughts. The SPI and SPEI findings align with historical drought reports, validating the indices as reliable monitoring tools. However, the divergence between SPI and SPEI during specific periods such as 2016 highlights the importance of including temperature-driven evapotranspiration in drought assessment.

The 3–7 day lag between rainfall and water balance changes presents a strategic window for intervention. For example, emergency alerts could be issued shortly after rainfall deficits to mitigate the impact on agriculture. These results align with Adikari et al. (2021), who found similar lag dynamics in South Asian flood systems.

5. Conclusion

This exploratory data analysis has revealed critical insights into the climatic drivers of floods and droughts in Bali LGA, Taraba State. It underscores the value of EDA in low-data environments, highlighting key correlations and temporal dynamics among climate variables. The study provides empirical support for incorporating derived indices such as SPI and SPEI in early-warning frameworks. Future research should build on this work by integrating satellite-based remote sensing, machine learning models, and real-time sensor data. A hybrid approach could significantly improve the accuracy and lead time of flood and drought forecasts. Policymakers and stakeholders should consider localized EDA insights when designing adaptation and mitigation strategies.

References

Abubakar, B., Umar, H., Barde, M., & Adamu, S. (2020). Socio-economic impact of flooding on the riverine communities of River Benue in Adamawa State, Nigeria. *[Journal/Publisher not provided – add if available]*.

Abubakar, B., & Ibrahim, M. (2016). Assessment of boreholes water quality in Bali Local Government, Taraba State, Nigeria. *FUTY Journal of the Environment*, 10(1), 1–12.

Adikari, Y., [add co-authors if any]. (2021). Predictive modeling of flood response using lags. *Environmental Modelling & Software*, 144, 105149.

<https://doi.org/10.1016/j.envsoft.2021.105149>

Alehile, K., & Alehile, K. (2023). Climate change effects on employment in Nigeria's agricultural sector. *Chinese Journal of Urban and Environmental Studies*, 11(2), 2350018. <https://doi.org/10.1142/S2345748123500185>

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts. <https://otexts.com/fpp2/>

Ibrahim, S. I., [add co-authors if any]. (2021). Penman–Monteith PET estimation. *Journal of Hydrology*, 598, 126413. <https://doi.org/10.1016/j.jhydrol.2021.126413>

Intergovernmental Panel on Climate Change (IPCC). (2022). *Climate change 2022: Impacts, adaptation, and vulnerability*. Cambridge University Press.

<https://www.ipcc.ch/report/ar6/wg2/>

Meteomatics AG. (n.d.). *Meteomatics: The global leader in weather intelligence*.

<https://www.meteomatics.com/>

Mosavi, A., Ozturk, P., & Chau, K. W. (2018). Flood prediction using machine learning models. *Water Resources Management*, 32(9), 3031–3055.

<https://doi.org/10.1007/s11269-018-1994-5>

National Emergency Management Agency (NEMA). (2022). *Annual disaster impact report*. NEMA.

Sseguya, F., & Jun, K.-S. (2024). Drought quantification in Africa using remote sensing, Gaussian kernel, and machine learning. *Water*, 16(18), 2656.

<https://doi.org/10.3390/w16182656>

United Nations Office for Disaster Risk Reduction. (2022). *Global assessment report on disaster risk reduction*. UNDRR. <https://www.undrr.org/gar2022-our-world-risk>

Wilks, D. S. (2011). *Statistical methods in the atmospheric sciences* (3rd ed.). Elsevier. <https://doi.org/10.1016/B978-0-12-385022-5.00001-4>

World Meteorological Organization (WMO). (2023). *State of climate services 2023*. WMO. <https://public.wmo.int/en/resources/state-of-climate-services>

A SOULBOUND NON-FUNGIBLE TOKENS-BASED CERTIFICATE VERIFICATION SYSTEM FOR NIGERIAN UNIVERSITIES

A Soulbound Non-Fungible Tokens-Based Certificate Verification System for Nigerian Universities

Ladan, K. S.^{1*}, Olaniyi, O. M.², Alfa, A. A.³, Salihu, A. E.⁴

^{1&4}*Department of Information Systems and Technology, National Open University of Nigeria, Abuja, Nigeria.*

²*Department of Cyber Security, National Open University of Nigeria, Abuja, Nigeria.*

³*Department of Computer Science, Confluence University of Science and Technology, Osara, Nigeria.*

Abstract

Nigerian educational institutions are faced with a growing problem of academic certificate verification, especially among employers and admission givers. The conventional approaches are prone to fraud and ineffective because of centralized administration, which diminishes the integrity of academic credentials of holders. This paper addresses the problem of insecurity, low scalability, and low tamper-proof nature of available systems to enable institutions to generate verifiable certificates. It removes intermediary influences during authentication of certificates by third parties, other educational institutions, employers and third parties. Thus, the paper presents a decentralized certificate verification system called NFTcate built from soulbound non-fungible tokens (SBTs) leveraging Ethereum smart contracts and IPFS metadata storage making academic certificates immutable, non-transferable, and openly verifiable. The proposed system enables enlisted educational institutions to mint certificates by permanently binding them to wallet addresses of the students. The performance of the NFTcate system was assessed using usability, cost-efficiency, reliability and speed parameters. The outcomes indicated that NFTcate averagely mints certificates at 20.0 seconds and verifies them at 1.2 seconds with lesser cost

implications when compared to the existing approaches. On the smart contract utilised by the NFTcate system, all test cases were passed for soulbound behavior, access control, and metadata uniformity. The results demonstrated a forward-looking and viable option for certificate counterfeiting in Nigeria and its educational system. Aside from academic credentials, the NFTcate opens the door to wider applications especially in professional licensing, identity verification, and safe digital record keeping, and contributing to the mounting body of research on blockchain-grounded credentialing.

Keywords: Blockchain, Soulbound NFTs, Certificate Verification, IPFS, Ethereum, NFTcate

1.Introduction

The educational certificate counterfeiting remains insistent problem largely the attacking credibility and integrity educational institutions and employment platforms globally. According to new findings, the academic certificates forgery was nearing dangerous proportions for the employers, educational institutions, and people looking for honest credentials and skills (Rani et al., 2024). Consequent upon this, prevalent fraud had eaten deep into public trust fabric concerning academic credentials which produces a biased advantage for perpetrators of the illegitimate educational paths. Common security approaches like holograms and watermarks, have become ineffective against present-day sophisticated printing technologies capable of illegal duplications of sensitive documents (Motuz et al.,2024).

Most times, a carefully generated forged certificate is highly unidentifiable against unavailable original copy. Interestingly, the process for confirming the genuineness of

educational credentials with concrete evidence is time-consuming and painstaking (Tumati et al., 2024). To add to that, the use of single-point databases for minting certificates and degrees are vulnerable to compromises in the attempt to exercise certain degree of control, which is relatively less-effective for cross-border certification verification and confirmation in case international degrees and diplomas (Rustemi and Dalipi, 2024). The reasons could be attributed to gaps in information and literacy as opined by Barone and Herbaut (2024), and the dearth of an infallible universal verification platform. The quality standardization measures of available schemes mentioned by Motuz et al. (2024) cannot be fully achieved including statistical variation analysis and vector compliance models.

Blockchain and Non-Fungible Tokens (NFTs) are two most-promising technologies for complementing their weaknesses. NFTs offer a new method of safeguarding educational credentials through the intrinsic transparency and immutability properties of blockchain technology. Educational institutions can generate digital assets through certificates tokenization, which are virtually less likely to falsify since their tamper-proof and seamless verifiable. NFT-based architecture conforms with General Data Protection Regulation (GDPR) like security and privacy. Delgado-von-Eitzen et al. (2024) developed an issuance and validation of academic certificates system based on NFTs. Noteworthy improvements in security and efficiency that NFT-based certification provides underscores its potential to fully change certificate management systems in numerous businesses and fields (Kamaleshwaran, 2023).

Additionally, Aparna and Kesavamoorthy (2023) demonstrated that blockchain strongly verify certificates, using NFTs' functions to counter phony degrees. These

innovations could help to rebuild broken down trust in academic degrees and diplomas by addressing the sources of academic fraud. Nevertheless, the real obstacle remains, which is the urgent need for a universally scalable, interoperable, and commonly recognized credential verification system.

Tumati et al. (2024) introduced a blockchain-based digital certificate system referred to as Soulbound Token Certification (SBTCert). It increases the verification system's effectiveness while maintaining decentralization, security, confidentiality and authentication. SBTCert created a safe and reliable framework for students, staff, and educational institutions in effectually tackling cases of racketeering and confirmation of certificates entitlements.

Nikolić et al. (2022) built a blockchain-grounded school management system called as Eternal Digital Certificate. It maintained the data related to the summer school events, the partakers, and the certificates issuance. The Polygon Supernets blockchain framework stores the data. The certificates are personalized as NFTs for partakers after successfully passing the courses allocated to their wallets. The certificates issued with the NFTs were non-transferable or disposable across wallets other than owners.

2. Methodology

The proposed system was built minding the user experience like an open Verification Access Point, an institution's Certificate Issuance Portal, and an institutions' and students' Personal Dashboards. These properties of the proposed system design reduce the complexity of blockchain frameworks. React creates a user-friendly interface, and exchanges between blockchain and IPFS interactions were facilitated by TypeScript and Node.js at the backend component. Pinata offers distributed storage

for certificate metadata. Solidity programming environment for smart contracts implementation.

The basic architecture of NFTcate's uses the smart contract that leverages on Soulbound NFTs, which is the non-transferable digital certificates before linking to the public address of the students. It eliminates unlawful circulation and fabrication by ensuring that non-transferable certificates can be minted and issued by authorized educational institutions. The IPFS hash are stored on the NFTcate for the purpose of confirmation and verification of issued certificates. The smart contract logic is initially built algorithmically to ensure that metadata of certificates including student_name, programme, grade, and issue date, were committed to the IPFS platform. This link is made by ensuring that the metadata sent to the IPFS storage contains the no Personally Identifiable Information (PII) and a randomly generated UUID for the student from the database to protect the personal information on-chain of the certificate owner in line with the National Data Protection Act (NDPA) enacted on June 12, 2023. Off-chain student data is stored on a secured database and the institution requires an authenticated jwt token to access said data. The frontend component generates simple user interface needless of wallet operations directly, whereas the backend component serves as a mediator for administering uploads of the metadata to the IPFS as well as managing NFTs minting across the smart contract. NFTcate also handles revocation by having a revocation flag on the Smart Contract mapped to the token ID. When a token is checked for validity, it returns both data and the revocation status. This ensures that the original record of the certificate is still available in case of checks but also returns that the certificate was revoked for any verifier requesting validation on the system.

During the evaluation of the NFTcate, universality, testability, cost-effectiveness, and speed were measured. Minting and recovery transactions gas consumption on the Sepolia Ethereum Testnet were determined to ascertain cost-efficiency. Certificate minting and Metadata retrieval times serve as system speed measure. By testing the verification procedure on a variety of devices and browsers, universality was guaranteed, ensuring that users without any prior understanding of blockchain technology or cryptocurrency wallets may use it. To verify essential features like transfer prevention, minting restrictions, and precise metadata linking, testability was lastly thoroughly carried out using automated unit and integration tests on the Solidity smart contract using Hardhat, Mocha, and Chai in addition to manual frontend testing.

2.1 System Architecture

The architecture of the system was built with usability and immutability in mind.

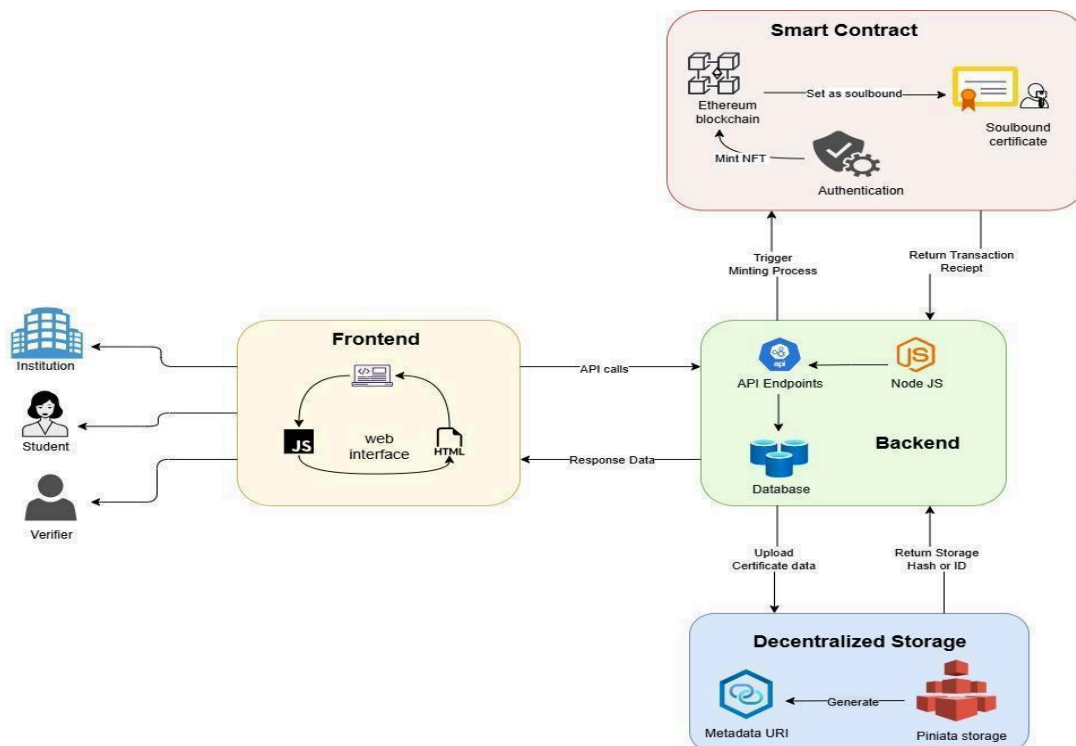


Figure 1: System Architecture. [Source: Authors' creation, draw.io]

NFTcate's architecture consists of 4 key components:

1. React-based frontend for user interactions
2. Node.js and TypeScript backend for handling smart contract and IPFS interactions
3. Solidity smart contracts deployed on the Ethereum testnet
4. IPFS via Pinata for decentralized metadata storage

The actors in the system as shown in Figure 1 are:

Institutions: These are the educational bodies responsible for issuing certificates. Institutions interact with the system to create and mint new certificates.

Students: These are the recipients of the certificates. These actors access their certificates through the system and can view and share the verifiable link.

Verifiers: These are individuals or entities (e.g., employers, other educational institutions) who need to verify the authenticity of a certificate.

The four key components of NFTcate are discussed as follows:

The Frontend web interface was built with React and Tailwind CSS, indicating a modern, responsive, and visually appealing user interface. This is the primary point of interaction for Institutions and Students. Institutions would use this interface to input metadata for certificates (e.g., student name, course, graduation date). Students would use it to view and potentially manage their issued certificates. The frontend web interface triggers functions on the backend via API calls then it gets response data and status once the triggered process has been completed.

This system leverages *decentralized storage* through the integration of IPFS (InterPlanetary File System) and Pinata. As a peer-to-peer network, IPFS makes it

possible to store and share hypermedia files in a distributed file system where files are recognized by their content rather than their location. By providing a simple tool to "pin" and manage files on IPFS, Pinata enhances this by ensuring their availability and permanence. This component's primary job is to store the actual certificate data and the metadata that goes with it in a decentralized, unchangeable manner. Here, certificate's details are saved upon issuance, creating a distinct IPFS hash for that set of information. The IPFS' hash is then transmitted to the backend for further handling and before passing it to blockchain integration in which the decentralized storage layer commits the metadata directly for storage from the frontend component.

The system's *backend* is a scalable and robust server-side platform developed using TypeScript and Node.js. The API endpoints are exposed to enable seamless communication of the system with the frontend component and the expected verifiers. This backend's interaction with decentralized storage is a key property because the system fetches the "IPFS Hash" from the Distributed Store as initial data to forge a direct link from the blockchain's NFT to the actual certificate data available on IPFS. The backend acts as a secure moderator for serious processes performed by the blockchain, which is introduced by this paper. Lastly, verifiers connect to the API endpoints of backend to recover certificate data linked to a unique NFT, upon successful confirmation of the NFT's availability and genuineness on the blockchain. The backend is essentially the core component of the system where all other components link to.

The development stage integrates a Solidity *Smart Contract* and is implemented on Sepolia Ethereum Testnet. The smart contract mints Soulbound NFTs which provides security for a certificate NFT that cannot be transferred or sold once it is assigned to

the wallet of a student as academic credentials. Furthermore, the smart contract's ownership integrity and immutability imposes non-transferability for all certificates. The system responds to "API Calls" through the frontend component to initiate the process of minting or querying of NFT data and then retrieves "Issue Certificate" instructions from the backend. Thereafter, the Soulbound NFT commences the minting process. Also, the verifiers query the blockchain smart contract directly for the purpose of determining availability of a certificate NFT's, metadata, and ownership as well as unique IPFS hashes. The smart contract also returns the transaction receipt after minting to be stored in the database by the backend component.

The smart contract's logic was outlined algorithmically as presented in figure 2. The algorithm considered the soulbound characteristics of certificates, institutional authority, and universality of verification.

The key design requirements that were considered were:

1. Only verified institution accounts can mint certificates.
2. Minted certificates are non-transferable (soulbound).
3. Each certificate must be linked to a student's public address (auto-generated or provided).
4. The metadata should include student name, institution, program, grade, issue date, and be stored on IPFS.
5. Anyone can verify a certificate using the token ID, which retrieves metadata from IPFS via the smart contract

BEGIN

IF caller IS NOT authorized institution

REJECT transaction

IF student address is invalid OR empty

REJECT transaction

PREPARE metadata

UPLOAD metadata to IPFS

STORE IPFS hash

CALL mintSoulboundNFT(student_address, token_id, ipfs_hash)

SET token transferability = false (non-transferable)

RETURN success

END

Figure 2: Pseudocode for the characteristic and behaviour of the soulbound NFT

3. Implementation

The NFTcate system effectively attains its main target of driving a decentralized method for verification of academic certificates through non-transferable (soulbound) NFTs and blockchain technology. The system design focuses on overcoming certificate racketeering and counterfeiting among Nigerian universities by generating verifiable, tamper-proof digital academic credentials, which are eternally allocated to the deserving student after their course of study.

NFTcate's backend is developed using TypeScript, Node.js, and a collection of API endpoints to manage key procedures. The API allows users to register either as students or institutions, with automatic generation of wallet addresses. Login functionality ensures secure authentication and returns relevant profile data while excluding sensitive information such as passwords. Institutions, once authenticated,

can mint certificates by sending metadata—including student name, course, grade, and date—to IPFS using Infura. The metadata is then used to mint a soulbound NFT, meaning it cannot be transferred or sold. There is also a verification endpoint that allows anyone to input a token ID and retrieve the original certificate data directly from IPFS and the blockchain.

The frontend, built with React, provides a smooth user experience tailored to both institutions and students. Users are introduced to the platform through a landing page that explains how NFT-based certificates work. The login and registration pages allow new users to create accounts securely. Once logged in, institutions gain access to a dashboard where certificate issuance is achieved by entering student data and uploading files. The minting process is accompanied by user feedback confirming successful operations. Students, on the other hand, can log in to their personal dashboards to view the certificates bound to their wallet addresses. These certificates are visually presented with data fetched in real time from IPFS. Additionally, there is a public verification page where any third party such as an employer can enter a token ID to confirm the authenticity of a certificate.

On the blockchain side as shown in Figure 3 below, NFTcate uses smart contracts written with OpenZeppelin's contract libraries and deployed on the Ethereum testnet (such as Sepolia). The contract restricts minting privileges through an authentication later to approved institution addresses and enforces non-transferability during the minting process, ensuring that once a certificate is issued, it remains permanently bound to the student's wallet. Each token's metadata is linked to IPFS by adding the token URI from the IPFS storage to the NFT, providing an immutable record of the certificate contents. The resulting hash is exposed to the student to be given to verifiers.

As shown in the snippet in Figure 3.5, The contract first checks if the institution is authorized before starting the minting process, and after the minting is complete the token is set as non-transferrable to seal the soulbound nature.

```
mapping(address => bool) public isInstitution;

modifier onlyInstitution() {
    require(isInstitution[msg.sender], "Not authorized");
    _;
}

function mintCertificate(address student, string memory
tokenURI) public onlyInstitution {
    uint256 tokenId = _tokenIdCounter.current();
    _safeMint(student, tokenId);
    _setTokenURI(tokenId, tokenURI);
    _tokenIdCounter.increment();
}

// Override to make NFT soulbound
function _beforeTokenTransfer(address from, address to, uint256
tokenId) internal override {
    require(from == address(0), "Soulbound: cannot transfer");
    super._beforeTokenTransfer(from, to, tokenId);
}
```

Figure 3.5: Snippet of solidity smart contract

A verifier requesting verification from the systems inputs the token URI which is checked by the smart contract for its existence.

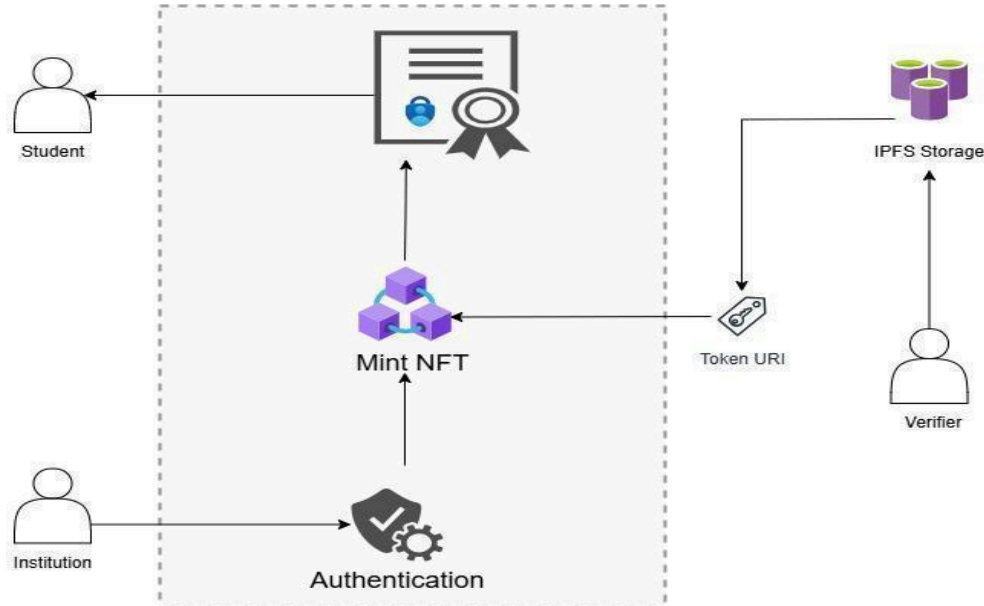


Figure 3: Smart contract certificate issuance flow [Source: Authors' creation, draw.io]

4. Results and Evaluation

Evaluation focused on aspects like cost-efficiency, speed and universality. To evaluate operational efficiency, gas costs for minting were recorded across multiple test transactions on the Sepolia testnet. Figure 4 shows a log of the cost of minting; it uses 171,511 gas at 0.022994941 gwei per gas on the Sepolia testnet. This brings us to a total of approximately 0.00000394 ETH for every minted certificate on the Sepolia testnet.

On the Mainnet, the price per gas usually goes for a range between 1 to 2 gwei with 1ETH being equal to 1 billion gwei. Given the total gas of 171,511, the average cost of minting a certificate on the Mainnet would be range from 171,511 to 343,022 gwei which

brings a total cost between 0.000171511ETH and 0.000343022ETH as depicted in Figure 3.

```
Gas Used: 171511
Gas Price: 0.022994941 gwei
Total Gas Cost: 0.000003943885325851 ETH
```

Figure 4. A log of the gas used, gas price and total cost on the Sopelia Testnet

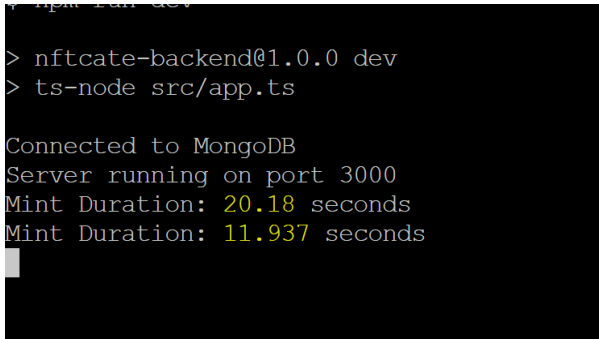
Ethereum occasionally experiences gas volatility during the periods of high demand because of how popular it has become but this can be tackled by scheduling issuance and minting periods during low volatility periods which can be easily tracked. In spite of these concerns, Ethereum remains the best option for high-value credentials due to its universality which ensures strong trust and guarantees and with its transition to Proof-of-Stake (PoS) consensus mechanism, its energy consumption has been reduced by over 90% lessening environmental concerns often associated with blockchain technology.

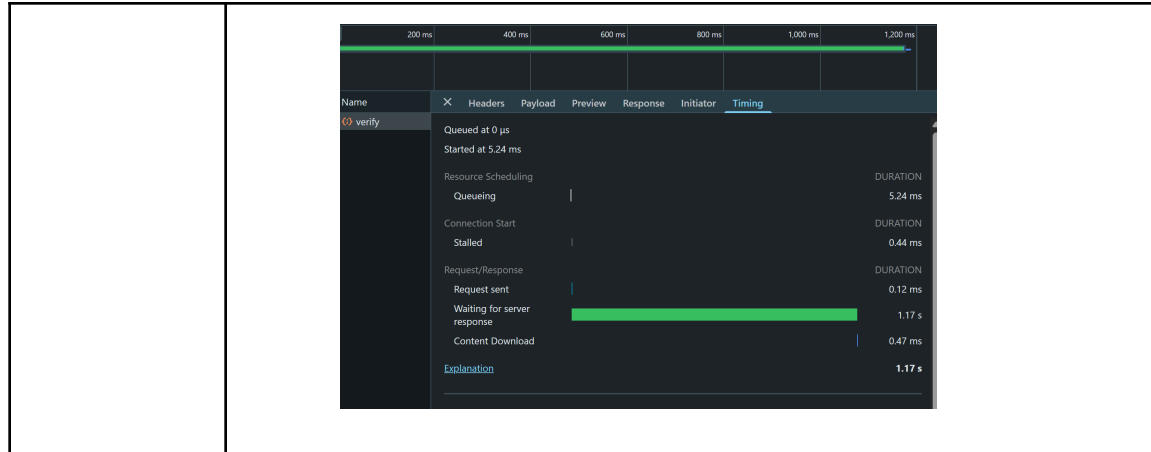
From Table 1, the performance timing was measured for key operations like the IPFS upload, Minting Transactions and Certificate Verification. The current desktop time was measured before the IPFS upload and Minting occurred and then the time on the desktop after the upload and minting operations. The start and end time were subtracted from each other to get the duration of each operation.

The process of uploading data to the IPFS on average took less than 1 second while minting an NFT took between 11 to 20 seconds depending on network stability and file size. Lastly, the verification process was tested by checking the browser networks tab

which showed that the total time from server waiting to response from the API call took about 1.2 seconds. This also would depend on browsing network stability.

Table 1: Performance table showing average duration of operations

Operation	Average Duration
IPFS upload	Less than 1 second  <pre> > nftcate-backend@1.0.0 dev > ts-node src/app.ts Connected to MongoDB Server running on port 3000 IPFS Upload Time: 0.605 seconds IPFS Upload Time: 0.62 seconds </pre>
Minting Transaction	Approximately 11-20 seconds depending on network stability and file size  <pre> > nftcate-backend@1.0.0 dev > ts-node src/app.ts Connected to MongoDB Server running on port 3000 Mint Duration: 20.18 seconds Mint Duration: 11.937 seconds </pre>
Certificate verification	Less than 1.2seconds



5. Conclusion

This research presented NFTcate, a decentralized certificate verification system that leverages the immutability and transparency of blockchain along with the non-transferable properties of soulbound tokens (SBTs) to address the persistent problem of certificate fraud in Nigerian universities. Dissimilar to the conventional approaches that are built on single-point verification documents or manual documents inspection, NFTcate generates and stores every certificate cryptographically. This secures and allocates uniquely to the deserving holders. These can be publicly verified needless of mediators. The proposed system enables only approved institutions to mint certificates through a mix of Ethereum smart contracts and distributed storage across the IPFS platform. Verifiers and owners (or students) and verifiers (employers, and other educational institutions) can readily view certificate metadata by means of exclusive token ID. NFTcate enhances accessibility by eliminating the need for cryptocurrency transactions or wallet knowledge since intricate blockchain activities are abstracted into a web-based interface. The system evaluation attained low operational costs, high reliability, and strong execution in realistic situations, which make NFTcate practical and scalable solution for nationwide adaptation. Comparatively, NFTcate addresses the weaknesses of soulbound

enforcement, minting centralization, and usability. It offers the strong foundation for larger applications in areas requiring tamper-proof and secure credentials minting and confirmation. Upcoming versions are expected to add offline authentication techniques like lightweight Progressive Web Apps (PWAs) and QR code with cached metadata to enhance accessibility, especially for low telecommunication and Internet penetration users. The system's basic architecture of immutable, secure, and non-transferable digital assets can be useful for minting employee credentials, voter IDs, and health data. With regards to improved cost efficiency, as it relates to the moments of market volatility of Ethereum, future works could look into adopting layer 2 scaling solutions (like Polygon and Arbitrum). This research focused on getting a proof-of-concept while showing cost efficiency, speed and universality. Though, more elaborate testing within actual scenarios are planned for future work, including looking into user and educational body surveys to compile a standardized format for student and certificate data for maximum uniformity while engaging willing Nigerian universities for real-life testing. Authors intend to evaluate the usability of the NFTcate system among students, verifiers, and institutions to determine the performance.

References

- Barone, C., & Herbaut, E. (2024). Decision-making mechanisms, information and educational inequalities. In *Research handbook on intergenerational inequality* (pp. 387–399). Edward Elgar Publishing. <https://doi.org/10.4337/9781800888265.00038>
- Delgado-von-Eitzen, C., Anido-Rifón, L., & Fernández-Iglesias, M. J. (2024). NFTs for the issuance and validation of academic information that complies with the GDPR. *Applied Sciences*, 14(2), 706. <https://doi.org/10.3390/app14020706>
- Edge Information. (2018, October 15). *Statistics: The problem of fake college degrees*. Medium.

<https://medium.com/@enterprisesolutionsforyou/statistics-the-problem-of-fake-college-degrees-2cc7e3a2093e>

Kamaleshwaran, B., Sneha, M., & Kavitha, S. (2023). Digital certification–Certification credential as non-fungible token (NFT). In *2023 2nd International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICAECA)* (pp. 1–7). IEEE. <https://doi.org/10.1109/ICAECA57943.2023.10234567> [add correct DOI if available]

Motuz, T., Dynovych, A., Babelchuk, O., Bykasova, S., & Gulevets, D. (2024). Examination of methods for assessing educational quality, the role of standardisation, and mechanisms for enhancing educational programme efficiency. *Multidisciplinary Reviews*, 7, 2024spe011. <https://doi.org/10.31893/multirev.2024spe011>

Nikolić, S., Matic, S., Čapko, D., Vukmirović, S., & Nedić, N. (2022). Development of a blockchain-based application for digital certificates in education. In *2022 30th Telecommunications Forum (TELFOR)* (pp. 1–4). IEEE. <https://doi.org/10.1109/TELFOR56148.2022.9983664>

Aparna, N., & Kesavamoorthy, R. (2023, November). A secure NFT-based system for eliminating counterfeit certificates. In *2023 IEEE 5th PhD Colloquium on Emerging Domain Innovation and Technology for Society (PhD EDITS)* (pp. 1–2). IEEE. <https://doi.org/10.1109/PHDEDITS60087.2023.10373697>

Rani, P., Sachan, R. K., & Kukreja, S. (2024). Educert-chain: A secure and notarized educational certificate authentication and verification system using permissioned blockchain. *Cluster Computing*, 27, 10169–10196. <https://doi.org/10.1007/s10586-024-04469-5>

Rustemi, A., & Dalipi, F. (2024, November). Academic certificate verification: A practical comparison between centralized and blockchain-based systems. In *2024 5th International Conference on Communications, Information, Electronic and Energy Systems (CIEES)* (pp. 1–6). IEEE. <https://doi.org/10.1109/CIEES62939.2024.10811200>

Tumati, T. V., Tian, Y., & Jiang, X. (2024). A soulbound token certificate verification system (SBTcert): Design and implementation. In *2024 IEEE 14th Annual Computing and Communication Workshop and Conference (CCWC)* (pp. 345–350). IEEE. <https://doi.org/10.1109/CCWC62308.2024.10456789>

**LEVERAGING GEOAI FOR MAPPING AND
FORECASTING OPEN DISTANCE LEARNING
EDUCATIONAL NEEDS AMONG YOUTH IN
NASARAWA STATE, NIGERIA**

Leveraging GeoAI for Mapping and Forecasting Open Distance Learning Educational Needs Among Youth in Nasarawa State, Nigeria

Christopher, C. V.

Abstract

Educational access in Nasarawa State faces significant constraints due to infrastructure deficits, teacher shortages, and insufficient investment in digital learning technologies. With Open and Distance Learning (ODL) expanding throughout Nigeria, particularly via institutions like Miva Open University and the National Open University of Nigeria (NOUN), understanding optimal deployment locations for digital educational resources becomes critical. This research examines how GeoAI, which merges Geographic Information Systems with Artificial Intelligence, can pinpoint underserved rural communities in Nasarawa State through satellite data analysis and demographic information. The study combines spatial analytics with educational and infrastructure metrics to identify educational gaps while proposing predictive models for future digital learning requirements. The findings, among others, revealed that when deploying educational solutions, providers should not focus on an LGA as a whole because, within an LGA, different educational needs exist. These results should assist policymakers, educational institutions, and ODL providers in targeting rural areas where investments in digital infrastructure, mobile learning technologies, or satellite-based classrooms could effectively address existing educational disparities.

1. Introduction

Nigeria has made considerable progress in educational access, yet substantial disparities persist between urban and rural populations. Remote communities frequently lack fundamental schooling infrastructure such as permanent buildings, qualified educators, transportation networks, electrical power, and internet access

(UNESCO, 2022). Nasarawa State's rural areas encounter additional obstacles from scattered populations and minimal government reach in isolated settlements (Abimbola et al., 2025).

Open and Distance Learning models have emerged as viable solutions for expanding educational reach. Digital platforms, mobile learning technologies, and self-directed study modules enable students in remote locations to access structured education without physical classroom attendance (Jegede, 2016). The National Open University of Nigeria has been instrumental in this transformation, developing comprehensive digital curricula and establishing regional learning centers (NOUN, 2020). However, scaling ODL in rural contexts requires spatial intelligence to inform planning and resource distribution (Tingzon et al., 2024).

Geospatial technology combines satellite imagery with demographic data to identify rural settlements with limited educational access and poor infrastructure (UN-GGIM, 2018). This study examines Nasarawa State as representative of rural Northern Nigeria, utilizing geospatial analysis to locate underserved communities and forecast future educational needs. This methodology can inform digital learning initiative prioritization, particularly in alignment with national educational strategies and development objectives (Federal Ministry of Education, 2020).

2. Problem Statement

Rural Nasarawa State communities experience an escalating educational crisis. Many settlements operate without functional schools, road connections, or power and internet infrastructure which are essential for digital education (NBS, 2020; Olaniran, 2021; Raufu, 2024). While government and non-governmental organizations have

implemented various educational interventions, these efforts are frequently hampered by insufficient monitoring systems, inadequate data, and inability to accurately identify areas of greatest need (Okoduwa & Amaechi, 2024).

Consequently, digital learning initiatives fail to reach the most vulnerable populations. Without clear spatial understanding of educational gaps, interventions may be misdirected, inadequate, or delayed. A geospatial approach is needed to map underserved rural settlements in Nasarawa State and model potential future digital learning demand (Gajardo et al., 2023). This would enable more strategic planning and effective resource allocation (UNESCO, 2022; UN-GGIM, 2018).

3. Research Aim and Objectives

This study aims to examine the application of GeoAI in Open & Distance Learning in rural Nasarawa State.

Specific objectives include:

1. Identifying and mapping educational need levels across Nasarawa State
2. Developing a novel framework for forecasting digital education needs
3. Providing recommendations for more effective targeting of rural education gaps by Open and Distance Learning providers

4. Literature Review

Gajardo et al. (2023) employed Sentinel-1 and Sentinel-2-time series data with Long Short-Term Memory deep learning architecture to create high-resolution Nigerian cropland maps. Their LSTM models showed substantial improvements in land use classification accuracy compared to traditional remote sensing products, particularly in heterogeneous landscapes. However, their work concentrated exclusively on agricultural zones without exploring how spatial outputs might inform social

infrastructure planning for education or healthcare. Additionally, the research focused on land classification without connecting to demographic data or service access.

Raufu et al. (2024) assessed vertical accuracy of various Digital Elevation Models (DEMs) including SRTM, ASTER, and AW3D using RMSE and MAE metrics against surveyed elevation points across different land covers in Ibadan. SRTM DEM proved more reliable in rural and semi-urban landscapes, while AW3D performed better in urban terrain. The study recommended DEMs for flood modeling, infrastructure planning, and spatial accessibility studies. While confirming elevation data usefulness, it didn't link findings to educational accessibility, population isolation, or digital readiness. The work emphasized model evaluation rather than practical deployment for social applications, though findings revealed effective modeling approaches for difficult terrain in digital education delivery, especially relevant to Nasarawa.

Using weakly supervised vision transformers and convolutional neural networks, Tingzon et al. (2024) identified school buildings in satellite imagery across ten African countries. Their approach required minimal labelled training data while producing highly accurate school location maps, achieving area under precision-recall curve scores exceeding 96%. However, the study didn't integrate these locations with infrastructure indicators like power supply, internet coverage, or population needs. It focused on physical school presence rather than modeling future digital learning demand or access gaps.

The reviewed literature demonstrates geospatial tools and satellite-based analysis power in education mapping, environmental monitoring, and land use classification across Africa. However, most studies exhibit significant limitations: limited infrastructure indicator integration (electricity, roads, mobile connectivity); narrow

focus on urban or environmental variables; lack of predictive modeling for Open and Distance Learning support; and minimal attention to underserved rural communities.

This research addresses these gaps by combining multiple geospatial datasets including road and electricity infrastructure alongside population density. The study contributes a replicable, data-driven framework for guiding digital education investments in Nasarawa and similar rural regions.

5. Methodology

Study Area & Tools

The analysis focused on Nasarawa State, Nigeria. All spatial processing and modelling were carried out in Google Earth Engine (GEE). Vector boundaries were drawn from FAO GAUL Level-1 (Nigeria → Nasarawa). Outputs were visualized in the GEE map interface and exported to QGIS for further analysis and reproducible figures.

Data Source

Table 1: Data Source and Type

Data Type	Purpose	Source	Format
Sentinel-2 imagery (10–20m resolution)	Detect built-up areas, vegetation, terrain in rural Nasarawa	Google Earth Engine	Raster (TIF)
WorldPop Settlement Layer	Population distribution at fine spatial resolution	WorldPop.org	GeoTIFF

Administrative boundaries	Nasarawa State and LGAs	GADM.org	Shapefile
Electricity Access	Assess digital device suitability	GRID3 Nigeria	Raster
Road Networks	Evaluate education resource accessibility	OpenStreetMap	Shapefile
Mobile network coverage (2G/3G/4G)	Assess online learning capability	GSMA Mobile	KML
Schools	Schools serve as input value for model training.	eHealthAfrica	GeoJSON

Pre-processing

All rasters were clipped to the Nasarawa boundary, Sentinel-2 scenes (2022) were filtered to

CLOUDY_PIXEL_PERCENTAGE < 20% and composited (median). While the electricity grid dataset (vector line shapefile) was converted into a distance raster using Euclidean distance in Google Earth Engine (GEE). This transformation allowed us to capture relative proximity of communities to the existing electricity network rather than treating grid lines as isolated features.

All datasets were projected to WGS 84 and aligned on a common grid to ensure spatial compatibility. Population and school location points were aggregated within the same grid cells, while accessibility features (roads, electricity, mobile signal) were expressed as continuous surfaces. These standardized layers were then combined through weighted overlay analysis in GEE to generate the education suitability index.

Suitability concept and feature engineering

We interpret education needs as a function of urbanity and accessibility: places that are built-up yet unlit and road-distant are more likely underserved for digital education. Accordingly, we engineered the following spatial features in GEE:

Settlement mask: GHSL built-up > 0 (settlement present).

Lighting class: VIIRS “well-lit” ($\text{avg\_rad} > 10$) vs unlit (≤ 10).

Road access class: within 2 km of a mapped road vs road-distant (> 2 km).

Spectral context: Sentinel-2 bands + NDVI + NDBI (to capture land-cover gradients).

Terrain: SRTM elevation (for broad physiographic differences).

Data-driven label construction

To avoid subjective drawing, we created labels directly from data using transparent rules: Low need (class 0): settlement present AND well-lit AND/OR within 2 km of a road.

Medium need (class 1): settlement present AND moderately constrained (e.g., lit but road-distant, or unlit but within 2 km of a road).

High need (class 2): settlement present AND unlit AND road-distant (> 2 km).

This approach reproduces a mask for “underserved”, interpreted as a built-up area that isn’t lit and 2km farther from the road. In addition, it extends to three classes to reflect a gradient of need.

We then converted this rule-based map into training samples using stratified sampling in GEE (balanced counts per class), ensuring labels came from objective proxies rather than hand-drawn polygons.

Model training and mapping

We trained a Random Forest classifier in GEE using the feature stack (Sentinel-2 bands, NDVI, NDBI, GHSL built density, VIIRS radiance, elevation). The model predicts three education-need classes (0/1/2). We also produced probability outputs from the classifier for uncertainty mapping.

The final map layers shown in GEE combined Nasarawa boundary (outline for context), TIGER roads and 2 km buffer (access), VIIRS night-lights (services/energy), GHSL built-up (settlements).

Sensitivity analysis

We re-ran the label construction with perturbed thresholds: night-lights splits at 8/10/12 and road buffers at 1/2/3 km. We tracked stability of class areas and validation scores. The model's conclusions are considered robust where performance and spatial patterns remain consistent across these settings.

Model Validation and Performance Metrics

To assess the reliability of the education suitability index, we applied both spatial and statistical validation techniques. First, we compared the predicted “high-need” zones with the actual distribution of existing schools from the eHealthAfrica database. Using grid cells as the unit of analysis, we computed accuracy, precision, and recall to evaluate how well the model distinguished underserved areas. Accuracy measured the overall agreement between predicted needs and observed school presence; precision captured the proportion of areas flagged as high-need that truly lacked schools; while recall measured the ability of the model to correctly identify underserved communities.

While we acknowledge that limited ground-truth education data restricts full-scale validation, this measure offers a practical performance check and provides confidence in the spatial patterns produced.

Map export and reproducibility

All layers were rendered in GEE for interactive inspection. The classified raster map was exported at 100 m resolution in EPSG:4326 to QGIS for further analysis and figure production. The script organizes layers in a consistent order (**boundary → roads → buffers → GHSL → night-lights → predicted need**), and includes a legend for classes.

Analysis and Interpretation

The geospatial analysis evaluated digital education needs across Nasarawa State by integrating geospatial indicators serving as proxies for educational exclusion and infrastructural marginalization. Analysis produced a predictive map classifying digital education needs across Nasarawa State using spatial indicators like nightlight intensity, settlement patterns, and road infrastructure proximity. Internal classifications derived from suitability indices were visualized through colour gradients: green areas indicate low need, yellow represents moderate need, and red signifies high digital education need.

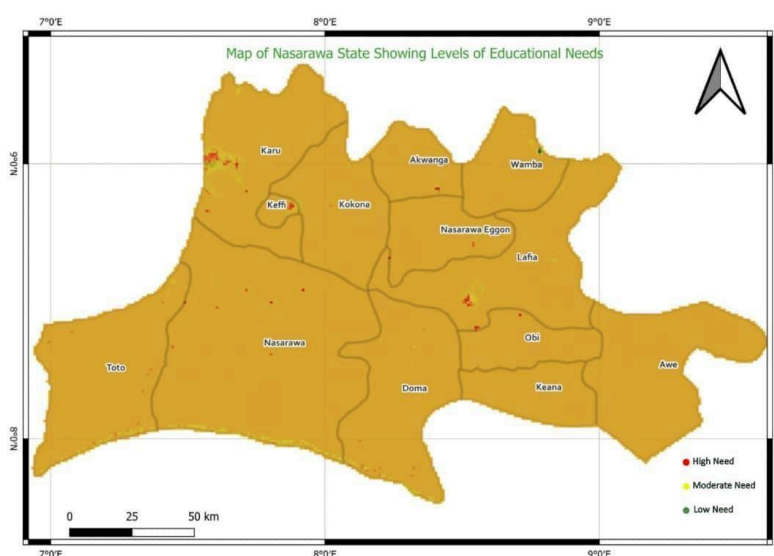


Fig 1: Map of Nasarawa State

A significant finding reveals substantial intra-LGA variation in digital education needs. Rather than each Local Government Area falling into single categories, many LGAs contain spatial mosaics of different need levels. For instance, Lafia and Karu, often considered urbanized and well-connected, exhibited green zones reflecting low need, especially within urban cores. However, these same LGAs showed substantial yellow and red patches in peri-urban and rural fringes, indicating that despite general infrastructural advantages, underserved clusters within these LGAs still face considerable educational exclusion.

In Lafia, while the central town exhibits strong nightlight intensity and dense settlements (suggesting better digital infrastructure), outskirts toward southern and eastern borders show weaker nightlight presence and greater primary road distances. These red-marked zones signal high digital learning needs, possibly due to poor electricity access and physical isolation. This dual pattern implies educational

interventions in Lafia must adopt localized targeting approaches rather than treating entire LGAs uniformly.

Similarly, Karu LGA, located near the Federal Capital Territory and influenced by urban development, displays green areas in well-lit, built-up towns. However, interior sections, particularly those distant from highway corridors, reflect higher digital exclusion levels with moderate to high needs. These findings suggest Karu's digital inclusion is spatially uneven and peripheral communities still require targeted Open and Distance Learning strategies.

LGAs including Doma, Toto, Nasarawa, and parts of Obi show more consistent red area distribution across their territories. These LGAs have minimal nightlight intensity, widespread built-up sparsity, and are often located far from major roads. Their classification as predominantly high-need zones is more uniform and clearly supports prioritization for digital education intervention.

The map reveals that some LGAs with moderate overall infrastructure, such as Nasarawa Eggon and Kokona, display mixed need levels, predominantly yellow with interspersed red and green patches. This implies transitional zones where some communities may already be receptive to blended or low-bandwidth digital learning, while others still require foundational infrastructure support.

The geographic patterns we found really shake up how we usually think about development planning. Typically, we look at entire Local Government Areas and assume they're pretty much the same throughout but that's not what we're seeing here. There's huge variation within single LGAs, which means policymakers can't just paint them all with the same brush. This is especially crucial when we're talking about

distance learning programs. These initiatives have to work with what's actually on the ground; whether there's reliable electricity, if people can get their hands on devices, how the transport networks function. If we just roll out the same approach across an entire LGA, we'll miss the isolated communities that might be struggling even when they're surrounded by areas doing relatively well.

What this research gives us is something we haven't had before; we can see the big picture across the whole state, but we can also zoom in and understand what's happening in specific places. From a practical standpoint, this changes everything about how we should approach digital education programs. Instead of broad-stroke solutions, we need targeted interventions, maybe solar-powered learning stations in some spots, mobile servers that work offline in others, or community radio programs where that makes sense. The urban areas are interesting because they're not just recipients of help; they can actually serve as launching pads for getting resources and content out to the rural communities that need them most.

This nuanced approach enhances digital education planning effectiveness and equity in Nasarawa State, ensuring no cluster is left behind due to misleading averages at higher administrative levels.

Discussion

This study provides evidence supporting obvious rural education and digital infrastructure gaps in Nasarawa State. The GeoAI model, combining satellite data (nightlights as electricity and connectivity proxy), road access, and settlement density, confirms longstanding qualitative observations with spatial and quantitative insights. Areas in Nasarawa and parts of Awe and Nasarawa Eggon consistently

showed low suitability scores, indicating poor digital access and infrastructure, confirming reports of rural marginalization in education (Adededeji & Olaniyan, 2011).

Findings reveal digital exclusion exists not only between LGAs but within them. While Karu and Keffi have higher overall access levels, some peripheral communities within these LGAs still face significant barriers. This internal disparity demonstrates the need for intra-LGA planning rather than relying on aggregate statistics for resource allocation.

The findings emphasize that conventional infrastructure development alone cannot close education gaps, but combining spatial insights with targeted interventions offers more effective strategies especially within a LGA.

Importantly, this model can be scaled nationwide. Because it relies on open-source tools like Google Earth Engine and QGIS, plus publicly available satellite data, other Nigerian states can replicate the approach at minimal cost. By bridging qualitative insights with spatial data, this method aligns with Nigeria's shift toward data-informed planning in digital and AI-led education (Federal Ministry of Education, 2020).

Comparison with Existing Frameworks

While our suitability model integrates geospatial datasets such as school locations, GHSL population distribution, TIGER road access, and VIIRS night-light intensity to assess education needs, it is important to situate this work against already existing initiatives in Northern Nigeria. One relevant framework is the Open and Distance Learning (ODL) program of the National Open University of Nigeria (NOUN), which has been piloted across several Northern states, including Nasarawa. The ODL framework

aims to expand access to higher education by establishing study centres in both urban and semi-urban locations, thereby reducing barriers related to distance and limited school infrastructure.

By comparing our geospatial suitability model with the ODL framework, we note a complementary relationship. Whereas the NOUN approach is institution-driven and focuses on higher education delivery through distance learning, our model provides an evidence-based, spatially explicit method for identifying areas of highest educational need (low, medium, high) based on accessibility, settlement type, and socio-economic proxies. This distinction highlights the novelty of our study: while NOUN targets expansion through infrastructure and digital delivery, our approach enables predictive mapping to guide where such interventions would be most impactful.

Implications for Open and Distance Learning (ODL)

The observed results from this study offer direct use cases for planning ODL interventions. First, underserved communities identified by the model can be prioritized and used for piloting. This includes rural Nasarawa and parts of Nasarawa Eggon, where conventional school access is limited and digital resources are lacking. Second, other similar studies provide frameworks for deciding where to deploy satellite internet services, mobile classrooms, or solar-powered ICT labs. Such technologies have proven useful in countries with similar access challenges (Shankar et al., 2021).

Finally, the deployment of teachers or the distribution of ICT resources should henceforth be carried out spatially. The model's classification reveals to education authorities to assign digital teachers to high-need areas and distribute tablets and laptops where most required.

Implementation Roadmap

For this framework to move from research into real-world application, it should be approached in phases. The first step is for key agencies such as the Ministries of Education and the National Bureau of Statistics to bring together the different strands of data such as satellite imagery, survey results, and existing administrative records, into a shared repository. Alongside this, training programs can equip education officers and local planners with the skills to use accessible cloud platforms like Google Earth Engine, which reduce the need for heavy infrastructure.

Once the foundation is in place, a pilot phase can begin in a few LGAs of Nasarawa state where rural communities face the greatest barriers. Here, the framework can be used to identify where new schools, mobile classrooms, or digital learning hubs would have the most impact. The results of this pilot such as improvements in enrolment or reductions in travel distance will help refine and strengthen the model.

The final step is to scale the approach nationwide, embedding it into the regular planning cycles of bodies such as the Universal Basic Education Commission. At this stage, results can be made accessible through interactive dashboards for government, NGOs, and donors, creating a transparent and adaptive system that evolves as new data becomes available. In this way, the framework becomes more than an academic contribution; it becomes a practical decision-support tool that advances Nigeria's education goals and the global commitment to inclusive and equitable learning.

Conclusion

This research demonstrates how GeoAI can effectively identify underserved areas for digital education in Nasarawa State. By analysing proxies for electricity, accessibility, and settlement presence, the model offers evidence-based foundations for planning ODL and digital inclusion while a predictive model handles the education needs as they evolve. The framework is developed with open source tools and resources such as GEE, QGIS, Satellite imageries, socio-economic datasets. Therefore, it is cost-effective, scalable, and adaptable across states. It also aligns with Nigeria's digital education and AI strategy as well as the SDG goal 4, providing smart planning tools for policymakers, NGOs, and education providers.

Once piloted in Nasarawa state, this model can support broader efforts to close digital gaps in education, especially in rural or semi-urban areas in the entire Nigeria.

References

- Adedeji, S. O., & Olaniyan, O. (2011). Improving the conditions of teachers and teaching in rural schools across African countries. *UNESCO International Institute for Educational Planning (IIEP)*. <https://unesdoc.unesco.org/ark:/48223/pf0000215066>
- Federal Ministry of Education. (2020). *Nigeria's digital education roadmap*. Abuja, Nigeria: Government of Nigeria.
- Gajardo, J., Rufin, P., van der Linden, S., Knauer, K., Pacheco, G., & Hostert, P. (2023). Mapping cropland in Nigeria using Sentinel-1/-2 time series and LSTM deep learning. *Remote Sensing*, 15(4), 1148. <https://doi.org/10.3390/rs15041148>
- Google Earth Engine. (2024). *Developer guide*. Google. <https://developers.google.com/earth-engine>

Jegede, O. J. (2016). Open and distance learning practices in Nigerian higher education: A survey of challenges and opportunities. *Open Learning*, 31(1), 1–10.

<https://doi.org/10.1080/02680513.2015.1125809>

National Bureau of Statistics (NBS). (2020). *Statistical report on education and infrastructure gaps across Nigeria*. Abuja: NBS.

National Open University of Nigeria (NOUN). (2020). *Annual report and strategic roadmap 2020–2024*. Abuja: NOUN Press.

Olaniran, S. O. (2021). Challenges facing the implementation of digital learning in Nigeria's rural communities. *Journal of Educational Technology Systems*, 49(3), 389–406. <https://doi.org/10.1177/0047239520977231>

QGIS Development Team. (2023). *QGIS user guide*. QGIS. <https://docs.qgis.org>

Raufu, S. A., Aladejana, J. A., & Lawal, K. M. (2024). Vertical accuracy assessment of digital elevation models in heterogeneous terrain using ground control points in Ibadan, Nigeria. *Journal of Geospatial Engineering*, 20(1), 34–50.

<https://doi.org/10.1016/j.jogeoeng.2024.02.005>

Shankar, S., Dwivedi, P., & Vora, S. (2021). Mapping education access using satellite data: A review of global applications. *Remote Sensing in Education*, 13(5), 251–267.

<https://doi.org/10.3390/rs13050251>

Tingzon, I., Lobos, A., Ball, J., & Blumenstock, J. (2024). Mapping school locations across Africa using weakly supervised deep learning on satellite imagery. *Nature Communications*, 15, Article 1292. <https://doi.org/10.1038/s41467-024-39765-4>

United Nations Educational, Scientific and Cultural Organization (UNESCO). (2022). *Reimagining our futures together: A new social contract for education*. Paris: UNESCO.

United Nations Committee of Experts on Global Geospatial Information Management (UN-GGIM). (2018). *Integrated geospatial information framework: Part 1 – Overarching strategy*. United Nations. <https://ggim.un.org>

VIIRS Nighttime Lights Data. (2023). NOAA National Centers for Environmental Information. <https://ngdc.noaa.gov/eog/viirs>

ENTROPY WITH DECISION TREE-BASED ANOMALY DETECTION FOR MITIGATING OPENFLOW-BASED DENIAL OF SERVICE ATTACKS IN SOFTWARE-DEFINED NETWORKS

Entropy with Decision Tree-Based Anomaly Detection for Mitigating OpenFlow-Based Denial of Service Attacks in Software-Defined Networks

*¹Terrang, U. M., ²Adebayo, O. S., ³Okunade, O. A., ⁴Abdulazeez, L., & ⁵Auru, F.
(^{1,4}Information Systems and Technology, ²Cybersecurity, ^{3,5}Computer Science)
Department, Faculty of Computing, National Open University of Nigeria, Jabi Abuja
Corresponding author: waleadebayo@noun.edu.ng*

Abstract:

Denial of Service (DoS) attacks pose persistent threats to Software-Defined Networking (SDN) due to the centralized role of the controller. Prior approaches include entropy-based anomaly detection, machine learning classifiers, and rule-based filtering. Braga et al. (2010) pioneered entropy detection, but such methods often generate high false positives. Li et al. (2021) demonstrated the potential of deep learning for SDN attack detection, though at the cost of significant computational overhead. Shin et al. (2013) AVANT-GUARD rule-based framework improved scalability but increased switch resource usage. Kumar et al. (2018) introduced SAFETY, combining entropy and Support Vector Machines (Machine Learning) for early SYN flood mitigation, which achieved improved accuracy but was largely attack-specific. This research advances these efforts by proposing an Entropy-Decision Tree-based framework for detecting and mitigating DoS attacks in SDN. The entropy detector identifies anomalous traffic distributions, while the decision tree classifier validates suspicious flows using features such as packet rate, flow duration, and byte count. Implemented in a Mininet environment with Ryu and tested using the UNSW-NB15 dataset, the model achieved 96.3% accuracy, a **3.2%** false positive rate, and low detection latency (19.2ms) with lightweight resource

consumption. The results demonstrate broader applicability across multiple attack vectors compared to entropy-only, machine learning only, or SVM-based models.

Keywords: Anomaly Detection, Denial of Service (DoS), Decision Tree, Entropy, Machine Learning, Mitigation, OpenFlow, Software-Defined Networking (SDN)

I. Introduction

Software-Defined Networking (SDN) has emerged as a transformative networking paradigm by decoupling the control plane from the data plane, thereby offering programmability, centralized control, and simplified network management. However, this centralization also exposes the SDN controller to significant security risks, particularly Denial of Service (DoS) attacks. Such attacks exploit the controller's critical role by overwhelming it with malicious traffic, leading to degraded performance, increased latency, or even network paralysis.

Several approaches have been proposed to address this challenge. Braga et al. (2010) introduced an entropy-based anomaly detection mechanism, which was computationally lightweight but prone to false positives. Li et al. (2021) applied deep learning techniques, achieving strong classification accuracy but requiring substantial computational resources, limiting scalability in real-time scenarios. Shin et al. (2013) AVANT-GUARD framework employed rule-based filtering to improve scalability and resilience; however, this approach placed additional strain on switches and required frequent rule updates. Kumar et al. (2018) advanced the field with the SAFETY model, which integrated entropy with Support Vector Machines (SVM) for early detection of SYN flood attacks. While this hybrid approach improved accuracy and reduced false

alarms, its focus was narrow, primarily targeting SYN floods, leaving broader attack vectors insufficiently addressed.

This research builds upon these prior efforts by proposing an Entropy–Decision Tree-based framework that integrates entropy analysis with a machine learning classifier for broader detection coverage. By combining statistical traffic features with lightweight supervised learning, the framework aims to reduce false positives, lower computational overhead, and provide effective mitigation across multiple types of DoS attacks, thereby overcoming the limitations of entropy-only, deep learning, or SVM-focused solutions.

Motivation for a Hybrid approach in order to address this gap, this research proposes the Entropy–Decision Tree Model, a multilayered hybrid defense framework. Entropy-based detection provides a fast, lightweight statistical layer for anomaly flagging, while Decision Tree classification validates anomalies using flow-level features to reduce false positives. The addition of a dynamic rule optimizer ensures adaptive mitigation, updating switch rules in real time to block or rate-limit malicious flows. Unlike deep learning, this model is lightweight; unlike entropy-only, it reduces misclassifications; unlike queueing, it incorporates context-aware intelligence.

The remainder of this paper is structured as follows: Section II reviews related works, Section III details the methodology, Section IV presents results, Section V discusses findings, and Section VI concludes with benefits, limitations, and future directions.

2. Related Works

The first exposed SDN's vulnerability to table-miss attacks, proposing the AVANT-GUARD framework Shin et al. (2013) that used connection migration and

actuating triggers. Gude et al. (2015) pioneered entropy-based monitoring, showing that entropy can detect high-rate anomalies but fails for stealthy low-rate attacks. Kreutz et al. (2015) provided a taxonomy of SDN threats, including flow table exhaustion and controller bottlenecks, emphasizing the need for adaptive detection. Pham et al. (2022) applied entropy to packet-in headers for OpenFlow-specific DoS, improving responsiveness but requiring pre-training.

Wang et al. (2021) used Random Forest on CICIDS2017, achieving 96.5% accuracy but with high feature engineering costs. Mousavi & St-Hilaire (2019) applied SVMs to SYN floods, yielding 94% precision but high false positives. Ahmad et al. (2021) introduced clustering for low-rate DoS detection, showing potential in offline analysis but excessive computation in real time. Wang et al. (2018) introduced packet-in queue control to protect controllers but risked misclassifying legitimate traffic.

Li et al. (2021) proposed CNN-based detection on UNSW-NB15, achieving 98.2% accuracy but with high resource demands. Ouhssini et al. (2024) developed DeepDefend (CNN-LSTM-Transformer), surpassing 99% accuracy in cloud environments, but impractical for resource-limited SDN controllers.

Jafarian & Al-Shaer (2018) introduced OF-RHM (Random Host Mutation) as a moving target defense but faced synchronization challenges. Liyanage et al. (2021) advocated in-switch filtering with stateless rules, effective against repetitive traffic but weak against polymorphic attacks. Ibrahim et al. (2022) integrated Ethereum blockchain for SDN-IoT, enhancing resilience but with latency overhead. Yin et al. (2022) used federated learning for multi-domain SDN, addressing privacy but limiting real-time adaptability.

Entropy-based detection is fast but imprecise. ML models are accurate but require heavy feature engineering. Deep learning offers high accuracy but is resource-intensive. Blockchain ensures trust but sacrifices efficiency. No approach simultaneously addresses accuracy, adaptability, scalability, and lightweight operation.

In addition to standalone approaches, some research has explored combining entropy with machine learning for DoS detection. For instance, Kumar et al. (2018) proposed SAFETY, a framework that integrates entropy features with machine learning models to detect and mitigate SYN flooding attacks in SDN. Their results demonstrated improved accuracy and faster mitigation compared to entropy-only approaches. However, their reliance on a single type of attack limited generalization, leaving room for more adaptive frameworks like the entropy–decision tree model presented in this study

The proposed Entropy–Decision Tree Model fills this gap by integrating entropy detection with Decision Tree validation and dynamic rule optimization, striking a balance between speed, precision, and resource efficiency.

3. Methodology

A. Proposed Framework (Entropy–Decision Tree Model) Overview

The proposed model integrates entropy-based anomaly detection with a decision tree classifier to detect and mitigate DoS attacks in SDN.

B. Environment Setup

The implementation was carried out in Ubuntu 22.04 LTS using the Mininet emulator. The Ryu controller was extended with an entropy detector and decision tree classifier. SYN flooding and UDP flooding attacks were generated using Scapy and hping3, while scalability tests were conducted with ONOS controller in larger topologies.

C. Data Preprocessing

Traffic logs were collected from Mininet experiments. Preprocessing included removing duplicates, normalizing features, and selecting attributes relevant to SDN traffic, including packet_in_count, entropy, flow duration, and byte count. From these, 36,500 samples were obtained, with 70% for training, 20% for validation, and 10% for testing. Additionally, 3,500 samples were reserved for real-time validation.

D. Entropy-Based Detection

Shannon entropy was computed on selected features (source IPs, ports). A significant drop in entropy indicated abnormal uniformity, flagging potential DoS activity.

E. Decision Tree Classifier

Suspicious flows detected by entropy analysis were passed to a decision tree classifier trained on labeled attack and benign traffic. Features such as entropy score, packet rate, and flow duration were used. The decision tree helped reduce false positives by validating anomalies contextually.

F. Dynamic Rule Optimization

Validated malicious flows triggered the dynamic installation of drop or rate-limiting rules on SDN switches, ensuring malicious traffic was mitigated while preserving legitimate flows.

G. Entropy-Decision Tree Model Architecture

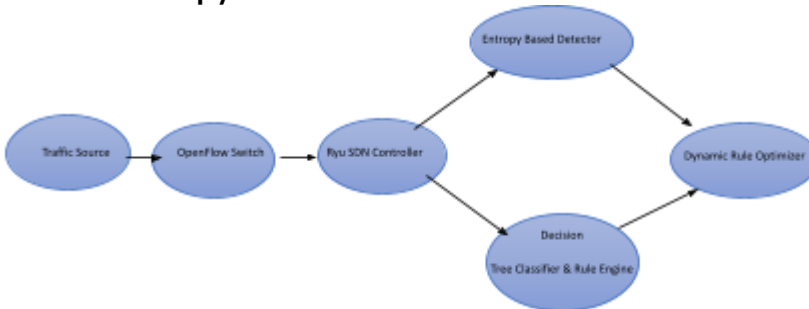


Figure 1.0: The Entropy-Decision Tree Model Diagram (Authors, 2025)

The Entropy-Decision Tree Model architecture presented in Figure 1.0 above includes the following components: Traffic Source, OpenFlow Switch, Ryu SDN Controller, Entropy-Based Detector, Decision Tree Classifier & Rule Engine, and Dynamic Rule Optimizer. The Traffic Source refers to all endpoints within or outside the network that initiate data transmission. These include legitimate users, services, IoT devices and malicious actors. In the context of Entropy-Decision Tree Model, the traffic source is monitored not for its content but for its behavior. Malicious traffic sources may attempt to flood the network with high volumes of forged or repetitive packets in order to overwhelm the SDN controller.

The OpenFlow Switch is the first network component that handles packets after they are transmitted by traffic sources. It operates at the data plane, forwarding packets based on flow rules installed by the SDN controller. If a packet does not match any existing flow rule, the switch issues a packet-in message to the controller for instructions on how to handle it. While this dynamic decision-making offers great flexibility and control, it also creates vulnerability. attackers can exploit it by generating vast numbers of unknown packets, thereby overwhelming the controller. In Entropy-Decision Tree Model, the switch serves a dual role: it detects unmatched packets and relays them to the controller and it also acts as a frontline defense once dynamic mitigation rules are installed. These rules, pushed by the Dynamic Rule Optimizer, can block or throttle suspicious traffic before it burdens the controller, making the switch both a traffic gateway and an enforcement node.

The Ryu SDN Controller serves as the central command hub of the Entropy-Decision Tree Model architecture. It controls the behavior of the entire network by making policy decisions and installing rules in the OpenFlow switch. What sets Entropy-Decision Tree Model apart is its extension of the Ryu controller with additional intelligence: the entropy-based detector and Decision Tree classifier are embedded as Ryu applications. The controller receives a continuous stream of packet-in messages, monitors flow-level statistics, computes entropy, evaluates behavior through Decision Tree models and then executes mitigation strategies. By acting as both a control center and an analytics engine, the Ryu controller in Entropy-Decision Tree Model is not merely reactive but capable of autonomous, intelligent decision-making. This positions it as the brain of the hybrid framework.

The Entropy-Based Detector is Entropy-Decision Tree Model's proactive surveillance system. It analyzes the randomness or unpredictability of specific traffic features, such as source IP addresses, MAC addresses or destination ports using Shannon entropy. In normal network operations, traffic is expected to show a high level of entropy, as it originates from a variety of users and devices. However, during a DoS attack, this diversity diminishes traffic becomes repetitive, predictable, and abnormally uniform. For instance, a SYN flood attack may result in hundreds of identical packets sent in a short time span, reducing entropy significantly. This module continuously computes entropy values over sliding time windows and compares them against pre-defined thresholds. If entropy falls below a critical value, it indicates the possibility of an attack and the traffic is marked for further classification. This early detection mechanism provides a crucial first layer of defense that allows Entropy-Decision Tree Model to catch anomalous behavior even before deeper analysis begins.

The Decision Tree Classifier and Rule Engine form the intelligent decision-making unit of Entropy-Decision Tree Model. After traffic has been flagged as suspicious by the entropy detector, it is passed to this component for a more refined analysis. Using a pre-trained Decision Tree model such as a decision tree classifier. The module examines several features including entropy, flow rate, packet size variation and inter-arrival time to determine whether the flow is malicious or benign. This dual-layered approach significantly reduces false positives by ensuring that not all anomalies are treated as threats.

The Dynamic Rule Optimizer is responsible for translating classification decisions into executable actions. It communicates directly with the OpenFlow switch to update or install flow rules based on current threat assessments. Its functions include dropping

malicious packets, rate-limiting suspicious flows and aggregating flow rules to prevent exhaustion of the switch's memory.

H. Implementation and Reproducibility Details

To ensure the reproducibility of the experiments, this section provides comprehensive details of the implementation, including hyperparameters, threshold selection methodology, and network topology.

Hyperparameter Tuning

The Decision Tree classifier was implemented using the Scikit-learn library in Python. The optimal hyperparameters were determined through a Grid Search with 5-fold cross-validation on the training set. The search aimed to maximize accuracy while minimizing model complexity to prevent overfitting and maintain low inference latency.

The final chosen parameters were:

- i. Criterion: Gini Impurity
- ii. Maximum Depth: 10
- iii. Minimum Samples Split: 5
- iv. Minimum Samples Leaf: 2
- v. Splitter: Best

These parameters provided the optimal trade-off between performance and computational efficiency for our specific task.

Entropy Threshold Selection

The entropy threshold is a critical parameter for triggering the secondary classification stage. Rather than using a fixed value, the threshold was dynamically derived from a statistical baseline of normal network traffic. During an initial training

phase, the mean (μ) and standard deviation (σ) of the entropy values for benign traffic were calculated over a 24-hour period. The detection threshold was then set to:

$$\text{Threshold} = \mu - 2\sigma$$

This approach ensures the system adapts to the natural variations of a specific network environment, flagging only traffic that deviates significantly from the established normal baseline.

Testbed Topology and Configuration

The emulation environment was built using Mininet (v2.3.0) to create a realistic SDN topology. The network consisted of a single Ryu controller (v4.34), 4 OpenFlow switches (Open vSwitch, v2.17.5), and 12 hosts. The links between nodes had a bandwidth of 100 Mbps and no artificial delay was added. This topology is representative of a small enterprise or data center network segment. The Ryu controller was extended with a custom application, which integrated the entropy detection and decision tree classification modules.

Software and Libraries

The experiment utilized the following key software versions to ensure consistency:

- i. Operating System: Ubuntu 22.04 LTS
- ii. Python: 3.10.12
- iii. Scikit-learn: 1.2.2
- iv. Ryu: 4.34
- v. Mininet: 2.3.0
- vi. Scapy: 2.5.0

I. Entropy-Decision Tree Model Algorithm

Begin

Continuously monitor incoming packet_in messages from OpenFlow switches

For each new packet_in:

Extract key flow features:

- Source IP / MAC address
- Packet rate
- Flow duration
- Byte count
- Source/Destination ports
- Packet size statistics

Append extracted data to observation window

If observation window duration expires:

Compute Shannon entropy of source addresses

If entropy < predefined threshold:

Mark current flow as suspicious

Construct feature vector:

- entropy, flow_duration, packet_rate, byte_count, src_port, dst_port,

packet_length_std

Use trained decision tree model to classify flow:

prediction $\leftarrow$ ML_model.predict(feature_vector)

If prediction = "DoS Attack":

Trigger mitigation:

- Install high-priority DROP rule for attacker IP
- Preserve or prioritize flows with legitimate patterns

Else:

Allow flow or install forwarding rule

Clear observation window and repeat monitoring

End

F. Evaluation Metrics The evaluation metrics used are Accuracy: Correctly classified traffic, False Positive Rate (FPR): Percentage of legitimate traffic misclassified, Latency: Detection-to-mitigation time CPU Utilization & Memory Consumption: Resource overhead on the controller and Scalability: Throughput across 10–100 switches and described below.

i. Accuracy (ACC)

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

TP (True Positive): Malicious traffic correctly classified as malicious.

TN (True Negative): Legitimate traffic correctly classified as legitimate.

FP (False Positive): Legitimate traffic incorrectly classified as malicious.

FN (False Negative): Malicious traffic incorrectly classified as legitimate.

Accuracy measures the overall correctness of the classifier. A higher accuracy means the framework is good at distinguishing between malicious and legitimate traffic.

ii. False Positive Rate (FPR)

$$FPR = \frac{FP}{EP+TN} \times 100\%$$

FP (False Positive): Legitimate flows misclassified as attacks.

TN (True Negative): Legitimate flows correctly classified.

FPR measures how often normal traffic is incorrectly flagged as malicious. A low FPR is crucial in SDN security because false alarms can lead to dropping legitimate traffic, disrupting network performance.

iii. Latency (L)

$$Latency = T_{detection} - T_{mitigation}$$

$T_{detection}$: Time at which attack traffic is detected.

$T_{mitigation}$: Time at which mitigation (e.g., rule installation) is applied.

Latency represents the delay between detecting an anomaly and applying countermeasures. Lower latency means the framework responds quickly, reducing the damage window of DoS attacks.

iv. CPU Utilization (CPU%) and Memory Consumption (MEM)

CPU Utilization:

$$CPU\% = \frac{CPU_{used}}{CPU_{total}} \times 100\%$$

Memory Consumption:

$$MEM = \frac{MEM_{used}}{MEM_{total}} \times 100\%$$

CPU_{used} , MEM_{used} : Amount of controller resources consumed by the detection modules.

CPU_{total} , MEM_{total} : Total available CPU and memory.

These metrics quantify the computational overhead introduced by the security framework. The goal is to maintain strong detection performance while keeping CPU and memory usage within acceptable limits so that the SDN controller remains responsive.

v. Scalability (S)

$$S = \frac{Throughput}{N_{switches}}$$

Throughput: Number of flows or packets processed per second.

$N_{switches}$: Number of SDN switches in the network (e.g., 10–100).

Scalability measures how well the framework handles increasing network size.

4. Result Presentation

Table 4.1: Experimental Result Summary

Framework / Authors	Accuracy (%)	FP	Latency (ms)	CPU (%)	Memory (MB)
Braga et al. (2010) Entropy only	90.5	7.8	15.6	12.4	98.2
Li et al. (2021) Deep Learning	97.8	3.9	28.5	26.3	152.7
Shin et al. (2013) Rule-based	92.7	5.1	21.4	19.6	134.4
Kumar et al. (2018) SAFETY (Entropy + SVM)	94.8	4.2	23.1	20.7	128.5
Proposed Framework (Entropy + Decision Tree)	96.3	3.2	19.2	18.1	121.5

5. Result Discussion

The comparative evaluation of the proposed Entropy–Decision Tree framework against existing detection models demonstrates a clear improvement in balancing accuracy, efficiency, and resource consumption. Braga et al. (2010) entropy-only approach, one of the earliest anomaly detection mechanisms for SDN, was computationally lightweight but suffered from significant false positives (7.8%). This was largely due to its reliance on a single statistical measure, which made it sensitive to normal traffic fluctuations that mimicked attack patterns. Although it achieved moderate accuracy (90.5%), its limited robustness undermined its practicality for modern, dynamic SDN environments.

Li et al. (2021) deep learning-based detection model improved significantly upon entropy-only methods, achieving an impressive 97.8% accuracy and reducing false positives to 3.9%. However, these gains came at the expense of scalability, as the model required substantial computational resources of 26.3% CPU utilization and 152.7 MB of memory and also incurred the highest latency among the compared methods at 28.5 ms. While suitable for offline training and environments with powerful infrastructure, such overheads make deep learning models less ideal for real-time SDN deployment where quick mitigation is critical.

Shin et al. (2013) AVANT-GUARD framework introduced rule-based filtering at the switch level, enhancing scalability and offering effective mitigation. It maintained reasonable accuracy (92.7%) and reduced latency to 21.4 ms compared to deep learning approaches. However, its reliance on frequent rule updates created overhead on the switches and introduced operational complexity. Furthermore, its false positive rate (5.1%) remained higher than advanced machine learning-based approaches, limiting its precision in distinguishing malicious from legitimate traffic.

Kumar et al. (2018) Kumar et al.'s SAFETY model marked a key advancement by integrating entropy features with Support Vector Machines (SVM) for early SYN flood detection. This hybrid approach demonstrated improved accuracy (94.8%) and reduced false positives (4.2%) compared to entropy-only methods. Nonetheless, the model's latency (23.1 ms) and memory consumption (128.5 MB) indicated non-trivial overhead, and its focus was narrow and mainly optimized for SYN floods. As a result, its adaptability to broader categories of DoS/DDoS attacks or zero-day scenarios was limited.

In contrast, the proposed Entropy–Decision Tree framework achieved the best trade-off across all metrics. With an accuracy of 96.3% and a low false positive rate of 3.2%, it strikes a balance between precision and computational efficiency. Its detection-to-mitigation latency of 19.2 ms was significantly lower than deep learning and SVM-based methods, and its resource consumption (18.1% CPU, 121.5 MB memory) remained lightweight, ensuring real-time deployability. Unlike Kumar et al. (2018) SAFETY, the proposed model was designed to detect a wider range of DoS/DDoS attack types beyond SYN floods, making it more versatile for dynamic and heterogeneous SDN environments.

Overall, these results highlight the strengths of the Entropy–Decision Tree model in addressing the shortcomings of prior works. It combines the speed and simplicity of entropy-based detection with the adaptability and classification precision of machine learning, while avoiding the heavy computational burden associated with deep learning or the narrow focus of SVM-based models. This positions it as a practical and scalable solution for securing SDN controllers against diverse denial-of-service threats.

6. Conclusion

This paper proposed an Entropy–Decision Tree hybrid framework for detecting and mitigating denial-of-service attacks in Software-Defined Networking (SDN). The approach combines the lightweight statistical power of entropy with the contextual intelligence of decision tree classification, resulting in improved accuracy, reduced false positive rates, and efficient use of controller resources. Compared with Braga et al. (2010) entropy-only method, the model demonstrates greater robustness against false alarms. It outperforms Li et al.'s deep learning system by delivering high detection

accuracy with lower latency and computational overhead, making it more practical for real-time deployment. Similarly, unlike Shin et al. (2013) AVANT-GUARD rule-based mechanism, it avoids excessive switch burden while ensuring precise detection. Furthermore, while Kumar et al. (2018) entropy-SVM model (SAFETY) shows early SYN flood detection capabilities, the proposed solution extends its effectiveness across broader DoS attack types, highlighting its adaptability to diverse threat vectors.

7. Future Works

Future extensions of this research will focus on enhancing adaptability against evolving attack patterns by incorporating unsupervised and reinforcement learning models capable of handling zero-day threats without extensive retraining. Additionally, large-scale validation in real-world and hardware-accelerated SDN environments will be pursued to complement simulation results. Finally, integrating federated learning and blockchain-based mechanisms into the framework offers promising directions for securing multi-domain SDN ecosystems while maintaining privacy, trust, and scalability.

References

- Ahmad, R., Wazirali, R., Bsoul, Q., Abu-Ain, T., & Abu-Ain, W. (2021). Feature-Selection and Mutual-Clustering Approaches to Improve DoS Detection and Maintain WSNs' Lifetime. *Sensors*, 21(14), 4821. <https://doi.org/10.3390/s21144821>
- Braga, R. , Mota, E. , & Passito, A. (2010). Lightweight DDoS flooding attack detection using NOX/OpenFlow. . *Proceedings of the IEEE Local Computer Networks Conference*, 408–415.

- Gude, N. , Koponen, T. , Pettit, J. , & Pfaff, B. C. M. , M. N. , & S. S. (2015). NOX: Towards an operating system for networks. *ACM SIGCOMM Computer Communication Review*, 38(3), 105–110.
- Ibrahim, N., Umer, M. F. , Ahmed, J. , & Jan, M. A. (2022). Blockchain-based secure communication for SDN-enabled IoT environment against DoS attacks. *IEEE Access*, 10, 24566–24577.
- Jafarian, J. H. , & Al-Shaer, E. ,. (2018). OpenFlow Random Host Mutation: Transparent Moving Target Defense using Software Defined Networking. *HotSDN '13: Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking*, , 127–132.
- Kreutz, D., Ramos, F. M. V., Esteves Verissimo, P., Esteve Rothenberg, C., Azodolmolky, S., & Uhlig, S. (2015). Software-Defined Networking: A Comprehensive Survey. *Proceedings of the IEEE*, 103(1), 14–76. <https://doi.org/10.1109/JPROC.2014.2371999>
- Kumar, P. , Tripathi, M. , & Nehra, A. , C. M. , & L. C. (2018). SAFETY: Early detection and mitigation of TCP SYN flood utilizing entropy in SDN. *IEEE Transactions on Network and Service Management*, 15(4), 1545–1559.
- Li, Y. , Xu, H. , & Zhao, S. (2021). A CNN-based detection model for DoS attacks in SDN using UNSW-NB15 dataset. *Future Internet*, 13(4), 94.
- Liyanage, M. , Braeken, A. , Ylianttila, M. , & Gurtov, A. (2021). Lightweight DoS mitigation for software-defined networks using switch-based filtering. *Computer Networks*, , 108231, 196.
- Mousavi, S. M. , & St-Hilaire, M. (2019). Early detection of DDoS attacks against SDN controllers. *IEEE International Conference on Computing, Networking and Communications (ICNC)*, 98–104.

- Ouhssini, A. , Azizi, M. , & Beni-Hssane, A. (2024). DeepDefend: A CNN-LSTM-Transformer hybrid deep learning framework for DDoS detection in cloud environments. *Future Generation Computer Systems*, 141, 135–149.
- Pham, Q. V. , Huynh-The, T. , Alazab, M. , & Le, L. B. , & H. W. J. (2022). Entropy-based anomaly detection for securing SDN against DoS attacks. *Future Generation Computer Systems*, , 129, 1–11.
- Shin, S. , Yegneswaran, V. , & Porras, P. , & G. G. (2013). AVANT-GUARD: Scalable and vigilant switch flow management in software-defined networks. fined networks. *ACM CCS*, 413–424.
- Wang, H. , Li, Z. , & Xu, Y. (2021). An SDN-based packet-in queue control mechanism against DoS attacks. *IEEE Access*, 9, 76650–76661.
- Wang, J. , Liu, Y. , & Feng, H. (2018). DDoS detection based on ensemble learning in SDN. *Journal of Physics: Conference Series*, 1069, 012076.
- Yin, Z. , Li, K. , & Bi, H. (2022). Trusted multi-domain DDoS detection based on federated learning. *Sensors*, 22(20), 7753.

**A REVIEW OF BEHAVIOURAL MACHINE
LEARNING MODELS FOR INTELLIGENT
DETECTION OF SOCIAL ENGINEERING
ATTACKS: ISSUES, TRENDS, AND RESEARCH
DIRECTIONS**

A Review of Behavioural Machine Learning Models for Intelligent Detection of Social Engineering Attacks: Issues, Trends, and Research Directions

¹Bashir, Z., ²Olaniyi, O. M., ³Kwanashie, A., ⁴Asekhome, A. A., & ⁵Nwoke, I. J.

^{1,3,4}Information Systems and Technology Department, ²Computer Science Department, ⁵CyberSecurity Department, National Open University of Nigeria, Abuja

Abstract

Social engineering attacks exploit human psychology to manipulate individuals into compromising systems, bypassing traditional technical defenses. Techniques such as phishing, spear-phishing, baiting, and pretexting have become increasingly sophisticated with the rise of digital platforms and personal data availability. Unlike malware or software exploits, these attacks target human weaknesses such as trust, fear, or urgency, rendering rule-based or signature-based detection mechanisms insufficient. Consequently, behavioral machine learning (ML) models have emerged to analyze user interaction patterns, communication behaviors, and linguistic cues for proactive threat detection. This article critically reviews behavioral ML models for social engineering detection, examining studies from 2020 to 2025 to identify issues, trends, and research directions. The period from 2020 to 2025 was selected as the focus of this review, because these years reflect the most recent advancements and studies in behavioral ML approaches for social engineering detection. Findings indicate that while supervised models such as Random Forests and Support Vector Machines are useful for behavioral classification, they have issues with class imbalance and scalability. Although they require a lot of data and processing power, deep learning models and natural language processing can pick up on tiny linguistic clues in phishing messages. Rule-based heuristics and machine learning are used in hybrid techniques that increase detection accuracy and decrease false positives, providing significant

promise for practical implementation. Nevertheless, issues including limited interpretability, antagonistic avoidance, and a lack of data continue to exist. While future directions emphasize the need for robust algorithms, wider applicability across IoT and decentralized systems, and ethical precautions, current research focuses on explainable AI, adaptive learning, and real-time detection. All things considered, behavioral machine learning models hold great potential for strengthening proactive and scalable defenses against changing social engineering threats.

Keywords: Social engineering, behavioral analysis, machine learning, phishing detection, anomaly detection, explainable AI, cybersecurity.

1. Introduction

Social engineering remains a formidable challenge in cybersecurity because it exploits psychological manipulation rather than technical vulnerabilities (Zaoui et al., 2024). Unlike malware or software exploits, social engineering attacks target human weaknesses such as trust, fear, or urgency to deceive individuals into compromising security systems (Akeiber, 2025). As cybercriminals continuously refine their tactics with phishing, spear phishing, baiting, and pretexting, traditional rule-based or signature-based detection mechanisms prove insufficient (Shehu & Bilalli, 2025). Consequently, the research community has turned to behavioral machine learning (ML) models capable of learning user interaction patterns, communication behaviors, and linguistic cues for proactive threat detection (Ugochukwu Ikechukwu Okoli et al., 2024).

This paper critically reviews behavioral ML models for social engineering detection, identifying issues, trends, and research directions. A systematic review approach was

adopted, screening relevant peer-reviewed studies published between 2020 and 2025 from databases including Scopus, Web of Science, and IEEE Xplore using keywords such as “social engineering detection,” “behavioral machine learning,” “cybersecurity,” and “AI-based threat detection.” The review first provides an overview of social engineering and ML applications in cybersecurity, then examines key ML techniques including supervised and deep learning models. The paper further explores hybrid approaches that integrate rules and ML, addresses challenges in behavioral detection, and concludes with emerging trends and future research pathways. The main research questions about behavioral machine learning (ML) for social engineering detection are covered in this article. These include which behavioral characteristics best predict attacks, how social engineering has changed in sophistication and tactics, and how a combination of technical and educational interventions can reduce cybersecurity risks that are focused on people.

Social Engineering and Machine Learning in Cybersecurity

Social engineering attacks denote a type of attacks (as shown in Figure 1) that exploit human vulnerabilities such as trust, cognitive biases, and emotional manipulation to bypass any traditional cybersecurity measures (Zaoui et al., 2024). Historically, social engineering has been a preferred method of attack among hackers, dating back to the 1970s. Unlike brute-force password cracking or software vulnerability exploits, social engineering relies on human error rather than computational skill. Attackers do not need to write complex code or penetrate firewalls; instead, they exploit trust and manipulate behaviors to gain unauthorized access. Since every digital system ultimately involves human interaction, these human factors are either overtly vulnerable or can be manipulated into becoming security weaknesses. As a result, social engineering constitutes a global cybersecurity risk (Wang et al., 2021). Social

engineering attacks primarily or mainly focus on psychological manipulation to deceive individuals into unknowingly diverting their sensitive information or performing actions that can compromise security (Akeiber, 2025). This method focuses on human pattern of thinking or social behavior such as the natural tendency to trust colleagues or authoritative figures, and to respond quickly to urgent requests. This makes social engineering very effective despite the presence of advanced technical defenses like firewalls and antivirus software. According to Alshammari et al., (2025), attackers often exploit cognitive biases like the scarcity effect, authority bias, or social proof to create convincing scenarios, thereby avoiding security scheme put in place by simply influencing human behavior.

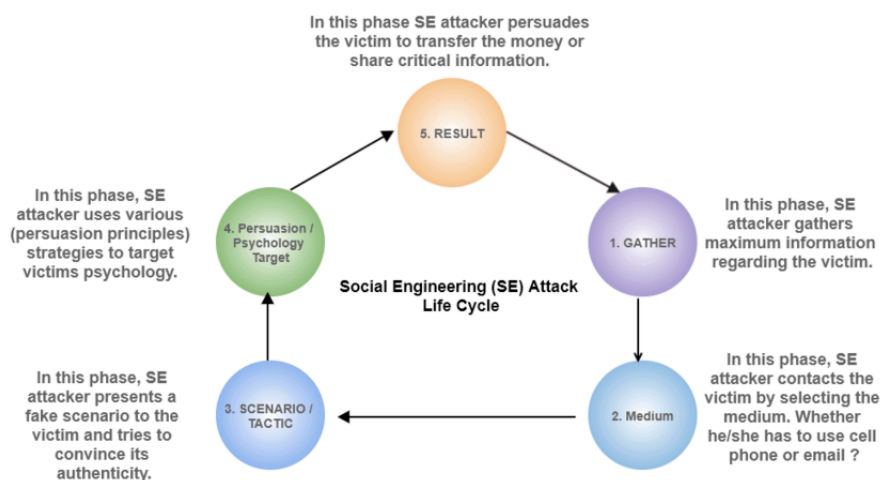


Figure 1: Social Engineering Attack LC (Yasin et al., 2021)

In recent years, the technicality and frequency of social engineering attack have increased exponentially, with phishing, spear phishing, baiting, and pretexting being some of the most prevalent forms (Shehu & Bilalli, 2024). Phishing approach involves indiscriminate mass messaging designed majorly to trick large numbers of users, while

spear phishing is highly targeted, often customized using personal information to increase credibility (Alshammari et al., 2025).

Moreover, baiting is another form of social engineering that entails enticing victims to unknowingly download malware or reveal information in exchange for a promised benefit, and creating a fake scenario to convince target victims to share confidential data (Aslam et al., 2024). This evolution in tactics has exposed the limitations of conventional signature-based detection methods, which rely primarily on recognizing known attack patterns or malware signatures. As a result, organizations are forced to create and adopt advanced detection and mitigation strategies that include behavior-based analytics, real-time anomaly detection, and enhanced user awareness training to effectively counter these human-centric threats (Okika, 2025).

The capability to leverage psychological manipulation rather than technical exploits makes social engineering attacks particularly challenging to counter. In social engineering attackers exploit both trust and various emotions such as fear, urgency, curiosity, and greed, pressuring victims into quick decisions that bypass rational security measures. Messages written to imply a false sense of urgency, can prompt unsuspecting recipients to click malicious link (Huseynov & Ozdenizci Kose, 2022; Akeiber, 2025). This human factor vulnerability is sometimes called "human hacking" (social engineering), which still remains critical threat up to today despite the continuously evolution of cybersecurity technology. In other to address these sophisticated means of attack, recent research are adopting multi-layered approach combining technological solutions, such as machine learning-based behavioral modeling, with ongoing employee education, strict access controls, and robust incident response procedures to build a resilient defense posture (Femi et al., 2025).

By examining particular aspects of human-computer interaction, behavioral machine learning models are able to make predictions. For instance, typing speed and keystroke dynamics can reflect stress or urgency reactions, whereas mouse movement patterns and trajectory variations can reveal hesitancy or job unfamiliarity. Anomalies that precede phishing success are frequently flagged by email metadata, such as sender patterns, timing, and link click behaviors. Because they immediately identify departures from a user's typical behavior which are frequently the first signs of a social engineering attempt before any technological signature is activated, features like these are more predictive.

Machine learning (ML) is one of the main branches of artificial intelligence use for developing intelligent system without explicit programming. ML is gaining popularity for its powerful capabilities in addressing the detection of social engineering attacks, which traditionally exploit human vulnerabilities rather than technical system (Okoli et al., 2024). However, unlike rule-based methods that rely on static signatures and predefined heuristics, behavioral ML models dynamically learn from large amount of datasets containing both communication and user patterns to identify traces of anomalies that can indicate present of attacks. This data-driven approach enables the models to detect anomalies from normal behavioral baselines, recognizing the detailed and evolving methods that social engineers use to craft messages and actions designed to fool people (Sri et al., 2025). Hence by continuously analyzing a wide scope of features such as email metadata, user interaction patterns, and communication content, ML systems offer real-time adaptive defense mechanisms capable of improving detection rates far beyond conventional techniques.

Furthermore, a key enabler of ML-driven social engineering detection is the integration of natural language processing (NLP) techniques, which improve models understanding and classification capability in term of semantic and syntactic pattern (Lansley et al., 2020). Social engineering attacks often hinge on carefully crafted deceptive language that manipulates recipients into taking harmful actions (Ashfi, 2024). Through NLP methods, ML algorithms can read and examine small language clues in emails, messages, and social media posts like strange phrases, mixed feelings, or unusual word choices that regular spam filters or blacklists might miss. This fusion of behavioral analytics with NLP equips detection systems to adaptively identify not only known social engineering schemas but also novel, previously unseen attack modalities, making machine learning a robust and scalable solution for safeguarding communication channels.

Moreover, modern ML models are designed to evolve continuously by leveraging ongoing input data streams, which is critical given that attackers perpetually refine their strategies to bypass defenses. Ensemble and hybrid learning approaches combine multiple algorithms to heighten robustness and accuracy, harnessing strengths from decision trees, random forests, support vector machines, and deep learning network (Martín et al., 2021). This adaptability coupled with large-scale data processing capabilities allows ML frameworks to operate effectively in diverse organizational environments, from enterprises handling millions of emails daily to critical infrastructure setups requiring instantaneous threat detection. Automated threat identification facilitated by ML significantly reduces manual workload on security analysts while providing scalable, proactive defense solutions that are essential to mitigating the increasingly sophisticated landscape of social engineering attacks (Okoli et al., 2024).

Behavioral Machine Learning Models

Behavioral ML models focus on analyzing user and communication behavior to flag suspicious activities. Multi-dimensional characteristics including keystroke dynamics, mouse movement patterns, login frequency distributions, email metadata, and conversational tone shifts are frequently extracted by behavioral machine learning models in order to accomplish this. Algorithms such as autoencoders for dimensionality reduction and anomaly detection or Hidden Markov Models (HMMs) for sequence analysis are then used to process these characteristics. An autoencoder, for instance, is able to learn a condensed representation of typical user behavior and identify notable differences as possible dangers. Likewise, ensemble methods balance recall and precision in high-variance settings by combining classifiers like Gradient Boosting Machines with anomaly detectors. These technological advancements enable behavioral machine learning algorithms to generalize toward zero-day or adaptable social engineering threats in addition to identifying previously observed attack vectors. (Pennada et al., 2025). Common approaches include supervised learning models like Support Vector Machines (SVM) and Random Forests, which classify behaviors based on labeled datasets of known attacks and legitimate interactions. These models excel in pattern recognition and have been effectively applied to phishing email detection, anomaly detection in network traffic, and user behavior analytics (Ayanbode et al., 2024; Pennada et al., 2025)

Building on these technical foundations, the subsequent subsections look at certain behavioral machine learning techniques, how they are used to identify social engineering assaults, and the issues that affect how effective they are in practical settings. An appraisal of recent papers is presented in Table 1.

Table 1: Meta-Analysis of Related papers

Title	Author(s)/Year	Objectives	Strength	Methodology	Limitation/Gap
AI-Driven System for Profiling Cybercrime Tendencies: A Preventive Approach	Edwin Onatuyeh, Duke Oghorodi, Greg Onwodi, Chukwuemeka Pascal Nwankwo, Emmanuel Ojei, Kingsley Ukaoha, Wilson Nwankwo 2024	To design a system that uses machine learning techniques to detect possible vulnerabilities to cybercrime by analyzing student behavioral, sociodemographic, and internet activity data.	Novel system predicting cybercrime tendencies using machine learning and behavioral analysis on survey datasets	Employed E-OOM with 1,050 anonymous online survey responses, expert validation, Python/Django with MySQL database	Dataset simulated or limited to specific groups (e.g., students), may not represent broader organizational behavior
AI-Driven Solutions for Social Engineering Attacks: Detection, Prevention, and Response	Hussam N. Fakhour, Basim Alhadidi, Khalil Omar, Sharif Naser Makhadmeh, Faten Hamad, Niveen Z. Halalsheh 2024	The objective of this study is to analyze social engineering attacks and evaluate AI-based countermeasures for identifying,	Innovative use of AI (ML, NLP, predictive analytics) to detect/counter social engineering; practical real-world	Case study-based demonstration of AI framework in modern social engineering attacks	Validation limited to certain datasets/environments; scalability and generalizability uncertain across diverse organizations

		stopping, and reacting to them.	approach with predictive capabilities		
Cyberattack Detection Framework Using Machine Learning and User Behavior analytics	Abdullah Alshehri, Nayeem Khan, Ali Alowayr, Mohammed Yahya Alghamdi 2022	The objective of this study is to provide a machine learning framework for detecting insider threats and other cyberattacks by utilizing recurrent neural networks and user behavior analytics.	Strong framework using RNN-LSTM and user behavior analytics; high accuracy (AUC=0.97)	Modeling user action sequences, neural network training; validated by evaluation metrics	Requires large, quality datasets; challenges in real-time deployment; possible false positives; scalability improvements needed
Using Machine Learning Algorithms to Predict Individuals' Tendency to Be Victim of Social Engineering Attacks	Farid Huseynov, Busra Ozdenizci Kose 2022	The study aims to use machine learning models to forecast a person's vulnerability to social engineering attacks.	Combines ML and behavioral data to predict vulnerability; uses simulated attack scenarios	Predictive modeling using participant responses to scenarios, personalized risk scores	Controlled simulations may not reflect real-world complexity; sample diversity limited; needs real-world testing and model refinement

			across industries		
A Survey for User Behavior Analysis Based on Machine Learning Techniques: Current Models and Applications	Alejandro G. Martin, Alberto Fernandez-Isabel, Isaac Martin de Diego, Marta Beltran	The study aims to evaluate, classify, and rank cutting-edge User Behavior Analysis works in order to direct future studies.	Comprehensive survey of 127 UBA studies, dual categorization, visual mapping of trends	Structured literature review with quantitative scoring and similarity metrics	No experimental validation or real-world implementation; potential literature bias; empirical testing needed
Social Engineering Attacks: Trends, Detection, and Prevention	Asere Gbenga Femi, Innocent L, Mangbon Jumbo Daniel	The study aims to examine social engineering assault patterns, detection techniques, and defense tactics.	Comprehensive overview integrating ML, NLP, behavioral analytics with technical and human-centered strategies	Thematic literature review of current tools and challenges	No empirical testing or case studies; lacks assessment of implementation in small organizations
Insider Threat Detection Using Behavioural	Pennada Siva Satya Prasad, Sasmita Kumari Nayak, M. Vamsi Krishna	The study aims to improve the detection of insider threats	Detailed feature categorization capturing	Classifier training and evaluation (Random Forest, ANN) on	Moderate performance in some categories; risk of overfitting; lacks real-time

Analysis through Machine Learning and Deep Learning Techniques	2025	through the use of machine learning models for feature categorization and SMOTE.	specific behavioral anomalies; uses SMOTE for class balance.	behavioral categories.	focus and cross-organizational generalizability.
Machine Learning in Cybersecurity: A Review of Threat Detection and Defense Mechanisms	Ugochukwu Ikechukwu Okoli, Ogugua Chimezie Obi, Adebunmi Okechukwu Adewusi, & Temitayo Oluwaseun Abrahams 2024	The study aims to examine how threat detection, protection strategies, and practical applications of machine learning improve cybersecurity.	Comprehensive overview of ML methods for cybersecurity including supervised, unsupervised, and reinforcement learning.	Analytical literature review of ML techniques and applications.	No empirical benchmarks or implementation frameworks; lacks real-time deployment evalon.
An Intelligent Cyber Security Phishing Detection System Using Deep	Ala Mughaid, Shadi AlZu'bi, Adnan Hnaif, Salah Taamneh, Asma Alnajjar & Esraa Abu Elsoud	The study aims to create and assess a machine learning model for precise	Comparative evaluation of ML algorithms for phishing detection	Supervised learning with feature engineering for email classification.	No adversarial robustness tests or real-world deployment; lacks behavioral/psychological factor analysis.

Learning Techniques	2022	phishing email identification.	with high accuracy on multiple datasets.		
Adversarial AI in Social Engineering Attacks: Large-Scale Detection and Automated Countermea sures	Anil Kumar Pakina, Deepak Kumar Kejriwal, Tejaskumar Dattatray Pujari 2025	The study aims to identify and stop AI-generated social engineering attacks, create a multi-modal AI security system.	Multi-moda l AI defense integrating transfer learning, deepfake detection, and behavioral anomaly systems.	Combines language models, audio analysis, and behavior detection, focused on financial security.	Resource intensive; requires large labeled datasets; limited real-world deployment; scalability and attacker adaptation unaddressed.

Behavioral machine learning (ML) models for detecting social engineering (SE) attacks have advanced considerably, leveraging user behavior data, predictive analytics, and AI techniques. Onatuyeh et al. (2024) proposed an AI-driven system to predict cybercrime tendencies using survey-based behavioral data, demonstrating the potential of combining machine learning with expert validation, though the dataset was limited to specific groups. Fakhouri et al. (2024) implemented an AI framework integrating ML, Natural Language Processing (NLP), and predictive analytics to detect social engineering attacks in practical scenarios, highlighting AI's capabilities while noting limitations in validation across diverse organizational contexts. Alshehri et al. (2023) utilized RNN-LSTM models with user behavior analytics to achieve high predictive accuracy, emphasizing the importance of modeling sequential user actions. Huseynov and Kose (2022) focused on predicting individuals' susceptibility to social

engineering through ML and behavioral data, whereas Martín et al. (2021) conducted a comprehensive survey of 127 studies on user behavior analysis, offering a structured overview of current models and applications. Femi et al. (2025) provided a thematic review integrating ML, NLP, and human-centered strategies, although empirical testing remains limited.

Collectively, these studies illustrate that behavioral ML models hold significant promise for identifying manipulative patterns, predicting potential victims, and informing preventive strategies. Nevertheless, challenges persist regarding dataset diversity, real-world deployment, scalability, and empirical validation, as summarized above in the meta-analysis Table.

Support Vector Machines (SVM) for Behavioral Classification

SVM is a supervised learning model widely used in behavioral cybersecurity due to its effectiveness in classification with high-dimensional data. In the context of social engineering, SVM models classify data or user actions into malicious or benign categories based on engineered behavioral features such as email metadata, interaction frequencies, and language usage patterns. However, by finding the optimal hyperplane that separates attack behaviors from normal behaviors in the feature space, SVMs can identify subtle deviations indicative of phishing or spear phishing attempts (Li, 2024). Their strength lies in handling linear and non-linear separations with kernel functions, but they require well-labeled training data and feature selection to perform optimally (Natha et al., 2025). Although SVMs are quite effective at identifying structured attack patterns, the requirement for high-quality, labeled datasets limits their efficacy. Because social engineering attacks are frequently infrequent, context-specific, and adaptive, it might be difficult to gather significant

amounts of behavioral data with appropriate labels. Because of this constraint, the model is inclined toward normal activity and causes issues like class imbalance, where benign activities often outnumber malevolent ones. Furthermore, when used on large, high-dimensional data streams produced in real-world networks, SVMs have scalability issues that lead to significant computing costs for both training and prediction. Lastly, unless SVMs are regularly retrained with current data, they run the danger of having poor generalization to novel or zero-day threats due to attackers' ongoing evolution of methods.

Random Forest and Decision Trees for anomaly detection of malicious user behaviors in social engineering

Random Forest is an ensemble of decision trees, providing robust behavioral detection by aggregating multiple decision paths to mitigate overfitting and improve accuracy (Heartfield et al., 2016). In the context of social engineering, study shows that these models are capable of analyzing user interaction patterns such as mouse movements, typing behaviors, and response delays, alongside email and network traffic features to detect anomalies that suggest manipulation attempts (Alt, Hassib, & Distler, 2023). They excel in interpretable results and handle mixed data types well, enabling adaptive learning from ongoing user behavior. Decision trees within Random Forest also help in pinpointing which behavioral traits most significantly contribute to detection, supporting explainability in threat identification (Pennada et al., 2025).

Deep Learning Models with Natural Language Processing (NLP) for Detecting Linguistic cues in Phishing and Social Engineering Attacks

Deep learning architectures particularly Recurrent Neural Networks (RNNs) and Convolutional Neural Networks (CNNs) are increasingly adopted to detect deceptive content in text by learning complex language patterns without explicit feature engineering. When combined with Natural Language Processing (NLP), these models process textual data from emails, chats, and social media to recognize phishing attempts by analyzing syntax, semantics, sentiment, and contextual cues inherent in social engineering messages (Femi et al., 2025). Deep learning's ability to capture long-range dependencies and subtle linguistic anomalies makes it highly effective for detecting tailored spear-phishing messages and evolving attack patterns, although it requires extensive training data and computational resources (Ashfi, 2024; Lee et al., 2020; Sri et al., 2025). In many text-based detection tasks, deep learning a subset of machine learning benefits from autonomous feature learning, which lessens the need for manual feature engineering.

Hybrid Behavioral Models for Social Engineering Detection

Hybrid machine learning models denote the fusion of rule-based heuristics with adaptive machine learning algorithms, designed to leverage the strengths of both paradigms in tackling complex cybersecurity challenges such as social engineering attacks (Cletus et al., 2022). The core principle behind hybrid models is layering: rule-based systems first filter communications or behaviors by applying expert-defined rules targeting known attack signatures or suspicious patterns. These rules operate efficiently by quickly discarding obvious benign or previously cataloged malicious instances, ensuring that only potentially ambiguous or unknown cases proceed to the next stage. This first filtering helps fix the problem that rule-based

systems are too strict and do not change easily. While they work well for known threats, they often struggle to keep up with new and changing attack methods (Awotunde et al., 2021).

Once suspicious inputs are flagged by the rule-based layer, machine learning algorithms take over, and different forms of hybridization emerge. Supervised models (e.g., Support Vector Machines, Random Forests) classify data or user activity into malicious or benign categories based on previously learned behavioral indicators. Unsupervised models, on the other hand, detect new or previously unseen attack patterns by recognizing deviations from established baselines of normal behavior. Deep learning models extend this adaptability by identifying subtle and complex anomalies that may be invisible to traditional ML, making them effective for detecting advanced spear-phishing or polymorphic attacks (Azam et al., 2023). These forms together allow hybrid systems to combine the precision of predefined rules with the flexibility of data-driven learning, significantly improving detection accuracy while simultaneously reducing false positives. Addressing false alarms is a vital benefit, as it helps security teams prioritize true threats, minimizing alert fatigue and improving operational efficiency (Rajabi & McArthur, 2024).

Moreover, hybrid models are advantageous for real-world cybersecurity deployment due to their scalability, interpretability, and continuous evolution capability. Scalable architectures can handle large volumes of data from heterogeneous sources such as emails, network traffic, endpoints, and user behavior logs, which is crucial in modern enterprise environments (Cletus et al., 2022). Interpretability stems primarily from the rule-based component, which provides clear and explainable decision criteria, making it easier for analysts to understand and validate automated alerts. At the same time,

the ML component adapts dynamically by integrating new threat intelligence and retraining on fresh data, enabling ongoing resilience to emerging social engineering tactics and polymorphic attacks. This evolutionary capability reduces dependence on manual rule updates, often a bottleneck in traditional systems, and supports proactive threat management. Generally, studies show that hybrid models provide a balanced approach where human expertise and machine intelligence are utilized to deliver robust, precise, and adaptive cyber defense (Hoang, 2023).

Table 2: Comparative Strengths and Limitations of ML Approaches for Social Engineering Detection

ML Approach	Application Area	Strengths	Limitations
Support Vector Machines (SVM)	Behavioral classification (language patterns, email metadata)	Manages high-dimensional data and is useful for detecting organized attacks.	Requires excellent labeled datasets; it has scalability problems and is prone to class imbalance.
Random Forest / Decision Trees	Anomaly detection in user behaviors (keystroke, delays, mouse)	Adaptive learning that is reliable, interpretable, and identifies important behavioral traits	Large-scale data and high processing costs; possible overfitting
Deep Learning (RNN/CNN + NLP)	Linguistic cue detection in phishing & spear phishing	Captures minute linguistic irregularities; useful for evolving assaults	Requires a lot of data, is computationally demanding, and has little explainability.
Hybrid Models	Combination of rule-based + ML	Minimizes false positives and strikes a compromise between accuracy and flexibility.	Deployment complexity, resource-intensive, and ongoing retraining

Challenges in Behavioral Detection for Social Engineering Attacks

Due to technological advancement several challenges affect the effectiveness of behavioral ML models. Several challenges affect the effectiveness of behavioral ML models for accurately detecting and preventing social engineering attacks. One significant issue is the high false positive rate, in which benign anomalies are mistakenly classified as attacks. This results in excessive false alerts and undermines confidence in detection systems. Furthermore, attackers continually adapt their strategies to evade identification by generating adversarial inputs designed to mislead ML classifiers. (Ayanbode et al., 2024; Akeiber, 2025).

The dependency on high-quality, comprehensive datasets is another obstacle. Training robust models requires diverse and representative samples of both malicious and legitimate behaviors. However, social engineering tactics are constantly evolving, and labeled data scarcity limits model generalizability across contexts and communications channels. Issues such as data bias, privacy concerns, and the need for real-time processing also complicate model deployment (Femi et al., 2025).

Explainability of ML models is a further concern. Security analysts require transparent insights into model decisions to trust automated threat assessments and to comply with regulatory standards. Black-box models, particularly deep learning ones, challenge interpretability, prompting a research focus on explainable AI techniques tailored for cybersecurity applications (Zhang et al., 2022).

Table 3: Summary of Challenges in Behavioral ML for Social Engineering Detection

Challenge	Description	Impact
High False Positives	Benign anomalies misclassified as attacks	Alert fatigue, reduced system trust
Adversarial Evasion	Attackers generate adversarial inputs to fool models	Models bypassed; reduced reliability
Data Scarcity & Diversity	Limited labeled datasets; evolving tactics not captured	Poor generalization, bias risks
Explainability	Black-box deep models lack transparency	Analysts distrust; regulatory compliance issues
Real-Time Processing	Need for scalable, low-latency detection	Deployment barriers in enterprise/IoT systems

Trends and Advances in Behavioral Machine Learning Models for Social Engineering Detection

Current research trends focus on the integration of explainable AI with behavioral ML to improve trust and actionable insights. This includes developing algorithms that provide human-understandable rationales for their predictions, facilitating better analyst intervention and collaboration. Additionally, other approaches such as the use of continuous learning and adaptive models that update based on new threat intelligence, maintaining effective detection against emerging social engineering tactics. Incorporating contextual information such as user roles, communication histories, and organizational policies enhances behavioral model precision (Ayanbode et al., 2024; Cletus et al., 2022; Mane et al., 2024).

Additionally, real-time detection capabilities using streaming data analytics and integration with automated incident response systems represent key improvements, enabling swift action to quarantine or mitigate detected threats before they cause

damage. The use of AI-driven simulations and sandbox environments to analyze attachments and links also supplements behavioral detection with technical malware analysis (Alt et al., 2025).

Research Trends Analysis

There has been growing global interest in behavioral machine learning approaches for social engineering detection, with significant contributions from countries such as Germany, India, and Nigeria. The annual distribution shows that the sharp rise in publications from 2022 to 2025, especially after 2023, indicates that detection research became increasingly critical as deepfake and generative AI technologies emerged. This geographical and temporal dispersion underscores the active influence of the Global South and North on the region in response to changing cyber threats. Important research trends highlight the move away from static models and toward more interpretable, dynamic, and automation-ready methods. Explainable artificial intelligence, adaptive models, continuous learning, and real-time detection improved by streaming analytics and incident response integration are common themes. The overarching trend emphasizes the convergence of automated defense, explainability, and behavioral analytics with the goal of developing comprehensive systems that combine analyst oversight and machine intelligence to thwart ever-more-sophisticated social engineering attacks.

Future Research Directions

Future research directions focus mostly on mitigating model vulnerabilities to adversarial attacks by designing resilient algorithms and defensive mechanisms (& Panchal, 2024). Other researchers suggest expanding the scope of behavioral models to cover emerging social platforms, IoT devices, and decentralized systems like blockchain and the likes, thereby ensuring broader applicability and resilience in detecting social engineering attacks. However, addressing ethical considerations such as data privacy, algorithmic fairness, and minimizing surveillance risks is integral to sustainable AI-powered cybersecurity. Enhanced collaboration between human expertise and AI systems is essential, promoting hybrid defense strategies that combine machine precision with human intuition.

There is also a growing interest in deploying explainable, trustworthy AI embedded in cloud and edge environments to enable scalable, distributed social engineering detection frameworks adaptable to diverse organizational needs.

Conclusion

Social engineering remains one of the most resilient cybersecurity threats due to its focus on human vulnerabilities. While traditional defenses remain limited, behavioral ML models demonstrate strong potential in identifying manipulative patterns and deceptive communication. Supervised models, ensemble methods, and deep learning architectures have advanced detection capabilities, while hybrid systems improve precision and scalability. Nevertheless, challenges such as adversarial evasion, data scarcity, data diversity, and model interpretability hinder widespread adoption. Current research trends emphasize explainability, adaptive learning, and real-time detection.

This trajectory illustrates how human oversight and machine intelligence are increasingly convergent.

Looking forward, future research must focus on designing robust algorithms capable of withstanding adversarial attacks, extending behavioral detection into new technological ecosystems such as IoT, social platforms, and decentralized systems, and embedding explainable AI into distributed environments like cloud and edge computing. Collaboration between human expertise and machine intelligence will be essential to balance precision with intuition. By prioritizing resilience, scalability, and ethical safeguards, the cybersecurity community can develop sustainable AI-powered defenses that empower organizations to stay ahead of increasingly sophisticated social engineering threats.

References

Akeiber, H. J. (2025). The evolution of social engineering attacks: A cybersecurity engineering perspective. *Al Rafidain Journal of Engineering Sciences*, 3(1), 294–316.

<https://doi.org/10.61268/r9c49865>

Alshammari, S. S., Soh, B., & Li, A. (2025). Understanding social engineering victimisation on social networking sites: A comprehensive review of factors influencing user susceptibility to cyber-attacks. *Information*, 16(2), 153.

<https://doi.org/10.3390/info16020153>

Alshehri, A., Khan, N., Alowayr, A., & Alghamdi, M. Y. (2023). Cyberattack detection framework using machine learning and user behavior analytics. *Computer Systems Science and Engineering*, 44(2), 1679–1689.

<https://doi.org/10.32604/csse.2023.026526>

Alt, T., Ibisch, A., Meiser, C., Wilhelm, A., Zimmer, R., Ditz, J., Dresen, D., Droste, C., Karschau, J., Laus, F., Müller, O., Neu, M., Plaga, R., Plesch, C., Sennewald, B., Thaeren, T., Unverricht, K., & Waurick, S. (2025). Generative AI models: Opportunities and risks for industry and authorities (No. arXiv:2406.04734). *arXiv*.
<https://doi.org/10.48550/arXiv.2406.04734>

Alt, F., Hassib, M., & Distler, V. (2023). Human-centered behavioral and physiological security. In *Proceedings of the New Security Paradigms Workshop (NSPW '23)* (pp. 1–14). ACM. <https://doi.org/10.1145/3633500.3633504>

Aslam, M. M., Kalinaki, K., Tufail, A., Naim, A. G. H., Khan, M. Z., & Ali, S. (2024). Social engineering attacks in industrial Internet of Things and smart industry. In *Emerging threats and countermeasures in cybersecurity* (pp. 389–412). John Wiley & Sons.
<https://doi.org/10.1002/9781394230600.ch17>

Awotunde, J. B., Chakraborty, C., & Adeniyi, A. E. (2021). Intrusion detection in industrial Internet of Things network-based on deep learning model with rule-based feature selection. *Security and Communication Networks*, 2021, Article 7154587.
<https://doi.org/10.1155/2021/7154587>

Ayanbode, N., Abieba, O. A., Chukwurah, N., Ajayi, O. O., & Daraojimba, A. I. (2024). Human factors in fintech cybersecurity: Addressing insider threats and behavioral risks. *International Journal of Multidisciplinary Research and Growth Evaluation*, 5(1), 1350–1356. <https://doi.org/10.54660/IJMRGE.2024.5.1.1350-1356>

Azam, Z., Islam, Md. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning-based model analysis through decision tree. *IEEE Access*, 11, 80348–80391. <https://doi.org/10.1109/ACCESS.2023.3296444>

Bilalli, M., & Shehu, D. V. (2024). Analysis of social engineering attacks and ways to increase user awareness and prevent them (Master's thesis, South East European University). SEEU Repository.

https://repository.seeu.edu.mk/sites/thesis/ThesisSharedDocs/MA_130371.pdf

Cletus, A., Weyory, B., & Opoku, A. (2022). Improving social engineering awareness, training and education (SEATE) using a behavioral change model. *International Journal of Advanced Computer Science and Applications*, 13(5).

<https://doi.org/10.14569/IJACSA.2022.0130572>

Fakhouri, H. N., Alhadidi, B., Omar, K., Makhadmeh, S. N., Hamad, F., & Halalsheh, N. Z. (2024). AI-driven solutions for social engineering attacks: Detection, prevention, and response. In *Proceedings of the 2nd International Conference on Cyber Resilience (ICCR 2024)* (pp. 1–8). IEEE. <https://doi.org/10.1109/ICCR61006.2024.10533010>

Femi, A. G., Mangbon, I. L., & Daniel, J. (2025). Social engineering attacks: Trends, detection, and prevention. *International Journal of Trendy Research in Engineering and Technology*, 9(3), 39–44. <https://doi.org/10.54473/IJTRET.2025.9305>

Martín, A. G., Fernández-Isabel, A., Martín De Diego, I., & Beltrán, M. (2021). A survey for user behavior analysis based on machine learning techniques: Current models and applications. *Applied Intelligence*, 51(8), 6029–6055.

<https://doi.org/10.1007/s10489-020-02160-x>

Heartfield, R., Loukas, G., & Gan, D. (2016). You are probably not the weakest link: Towards practical prediction of susceptibility to semantic social engineering attacks. *IEEE Access*, 4, 6910–6928. <https://doi.org/10.1109/ACCESS.2016.2616285>

Hoang, N. V. (2023). Human expertise and machine learning in collaborative intelligence frameworks for robust cybersecurity solutions. *Journal of Applied*

Cybersecurity Analytics, Intelligence, and Decision-Making Systems, 13(12), 1–12.

<https://sciencespress.com/index.php/JACAIDMS/article/view/2023-12-04>

Huseynov, F., & Ozdenizci Kose, B. (2022). Using machine learning algorithms to predict individuals' tendency to be victim of social engineering attacks. *Information Development*, 40(2), 298–318. <https://doi.org/10.1177/02666669221116336>

Lansley, M., Mouton, F., Kapetanakis, S., & Polatidis, N. (2020). SEADer++: Social engineering attack detection in online environments using machine learning. *Journal of Information and Telecommunication*, 4(3), 346–362. <https://doi.org/10.1080/24751839.2020.1747001>

Lee, Y., Saxe, J., & Harang, R. (2020). CATBERT: Context-aware tiny BERT for detecting social engineering emails (No. arXiv:2010.03484). *arXiv*. <https://doi.org/10.48550/arXiv.2010.03484>

Li, P. (2024). Machine learning techniques for pattern recognition in high-dimensional data mining (No. arXiv:2412.15593). *arXiv*. <https://doi.org/10.48550/arXiv.2412.15593>

Mane, S., Kundu, S., & Sharma, R. (2024). A survey on online user aggression: Content detection and behavioral analysis on social media (No. arXiv:2311.09367). *arXiv*. <https://doi.org/10.48550/arXiv.2311.09367>

Mughaid, A., AlZu'bi, S., Hnaif, A., Taamneh, S., Alnajjar, A., & Abu Elsoud, E. (2022). An intelligent cyber security phishing detection system using deep learning techniques. *Cluster Computing*, 25, 3819–3828. <https://doi.org/10.1007/s10586-022-03604-4>

Natha, P., Tera, S. P., Chinthaginjala, R., Rab, S. O., Narasimhulu, C. V., & Kim, T. H. (2025). Boosting skin cancer diagnosis accuracy with ensemble approach. *Scientific Reports*, 15(1), 1290. <https://doi.org/10.1038/s41598-024-84864-5>

Okika, N. (2025). Mitigating insider threats and social engineering tactics in advanced persistent threat operations through behavioral analytics and cybersecurity training.

ResearchGate. <https://www.researchgate.net/publication/389464287>

Onatuyeh, E., Oghorodi, D., Onwodi, G., & Nwankwo, C. P. (2024, November). AI-driven system for profiling cybercrime tendencies: A preventive approach. In *2024 IEEE 5th International Conference on Electro-Computing Technologies for Humanity (NIGERCON)*. IEEE. <https://doi.org/10.1109/NIGERCON62786.2024.10927117>

Patel, P. S., & Panchal, P. (2024). Adversarial attacks on machine learning-based cyber security systems: A survey of techniques and defenses. *International Journal of Electronic Security and Digital Forensics*, 17(1-2), 1-?.
<https://doi.org/10.1504/IJESDF.2025.143481>

Pakina, A. K., Kejriwal, D. K., & Pujari, T. D. (2025). Adversarial AI in social engineering attacks: Large-scale detection and automated counter measures. *International Journal of Science and Technology*, 4(1). <https://doi.org/10.56127/ijst.v4i1.1964>

Pennada, S. S. P., Nayak, S. K., & M, V. K. (2025). Insider threat detection using behavioural analysis through machine learning and deep learning techniques. *International Research Journal of Multidisciplinary Technovation*, 74-86.
<https://doi.org/10.54392/irjmt2527>

Rajabi, F., & McArthur, J. J. (2024). Applying OPTICS with and without PCA for fault detection of fan coil units using building automation system data. *Energy and Buildings*, 317, 114368. <https://doi.org/10.1016/j.enbuild.2024.114368>

Sri, B. G., Sohan, M. D., Rao, N. S., Charishma, N. S., & Deepika, Y. A. (2025). Understanding phishing: Creating effective strategies for detection using Phishscan.

Journal of Cybersecurity Engineering, 3(2), 130–142.

https://jcse.cloud/JCSE/Published_Papers/V3I2S6P130_142.pdf

Okoli, U. I., Obi, O. C., Adewusi, A. O., & Abrahams, T. O. (2024). Machine learning in cybersecurity: A review. *World Journal of Advanced Research and Reviews*, 21(1), 2286–2295. <https://doi.org/10.30574/wjarr.2024.21.1.0315>

Yasin, A., Fatima, R., Liu, L., Wang, J., Ali, R., & Wei, Z. (2021). Understanding and deciphering of social engineering attack scenarios. *Security and Privacy*, 4(4), e161. <https://doi.org/10.1002/spy2.161>

Zaoui, M., Yousra, B., Yassine, S., Yassine, M., & Karim, O. (2024). A comprehensive taxonomy of social engineering attacks and defense mechanisms: Toward effective mitigation strategies. *IEEE Access*, 12, 72224–72241. <https://doi.org/10.1109/ACCESS.2024.3403197>

Zhang, Z., Hamadi, H. A., Damiani, E., Yeun, C. Y., & Taher, F. (2022). Explainable artificial intelligence applications in cyber security: State-of-the-art in research. *IEEE Access*, 10, 93104–93139. <https://doi.org/10.1109/ACCESS.2022.3204051>

COMPARATIVE ANALYSIS OF QUERY RETRIEVAL AND PERFORMANCE OF XML, SQLITE, AND FLAT FILE DATABASE

Comparative Analysis of Query Retrieval and Performance of Xml, Sqlite, and Flat File Database

Abubakar, A. A., Obunadike, G. N., & Yahaya, U.

National Open University of Nigeria

Corresponding Author: adamuahmad14@gmail.com

Abstract

The exponential increase in digital information has heightened the importance of database performance in modern computing. Traditional relational database management systems (RDBMS) have been dominant for decades, but they often exhibit bottlenecks when managing large or complex datasets. This study presents a comparative performance evaluation of three alternative data storage approaches—XML, SQLite, and Flat File databases—focusing on query retrieval, insertion efficiency, and scalability. Using datasets of 1,000, 5,000, and 10,000 records, converted into multiple formats via custom-built Java algorithms, retrieval and insertion times were measured under controlled conditions. Queries were optimized by eliminating middleware and employing direct execution: JDBC for SQLite, XPath for XML, and buffered I/O for Flat Files. Results indicate that Flat File databases are highly efficient for lightweight, small datasets, achieving retrieval times as low as 6.6ms for 1,000 records. However, scalability issues emerged at higher volumes, with performance deteriorating significantly beyond 5,000 records. SQLite demonstrated superior scalability, maintaining retrieval efficiency at 10,000 records with an average of 36.7ms. Conversely, XML databases consistently excelled in insertion tasks, recording 85.95ms for 10,000 records, though retrieval times remained the slowest across all tests. These findings suggest that the optimal database choice depends on workload characteristics: Flat Files for lightweight applications, SQLite for larger datasets, and XML for insertion-heavy use cases. The study contributes practical

benchmarks and query optimization strategies, providing actionable insights for software engineers, database administrators, and researchers designing data-intensive applications.

Keywords: Benchmarking, Database performance, Data storage Query, Flat File retrieval, SQLite, XML,

1.0 Introduction

Databases are any organized collections of data or information that a computer can quickly search through and retrieve. Databases are designed to make different data-processing operations—like storing, retrieving, modifying, and deleting data—easy. Information is retrieved from the database in response to queries by a database management system (DBMS) (Erik, 2023).

The modern information ecosystem is increasingly driven by data. The rapid expansion of big data, Internet of Things (IoT), and cloud-based applications has placed unprecedented demands on database performance. Efficient storage, retrieval, and insertion of data are critical not only in enterprise computing but also in personal, mobile, and embedded systems (Candel et al., 2022). Traditional relational database management systems (RDBMS) have been widely adopted for their strong consistency and transactional guarantees. However, their rigidity and resource requirements often result in performance bottlenecks, particularly when datasets grow in volume and complexity (Panwar, 2025).

Many databases hold information primarily in the form of natural language texts of documents; databases that are number-oriented, on the other hand, save information

largely in the form of financial data, tables, statistics, and unprocessed scientific and technical data. Individuals can utilize personal computer systems to keep small databases at home. An increasing number of data sets that were once distinct are being electronically merged into bigger collections called data warehouses. Subsequently, "data mining" software is utilized by businesses and government bodies to examine various parts of the data and identify trends (Christopher, 2022).

Alternative database models have emerged to address these limitations. Lightweight relational engines such as SQLite offer embedded functionality with minimal overhead, making them highly attractive for mobile applications, browsers, and IoT devices (Gaffney et al., 2022). Document-oriented storage models, such as XML databases, provide flexibility in representing hierarchical and semi-structured data but at the cost of slower retrieval performance (Twinkle et al., 2020). Flat File databases, though simple, remain relevant in applications requiring minimal setup and fast access to small-scale datasets (Fowler, 2023).

Despite their widespread use, there is a lack of detailed comparative studies evaluating XML, SQLite, and Flat File databases under similar experimental conditions. Most existing studies either focus exclusively on relational versus non-relational comparisons or highlight only one alternative in isolation. This research addresses that gap by empirically analyzing query retrieval and insertion performance across increasing dataset sizes.

1.2 Statement of the Problem

In the contemporary landscape, most data sources are primarily stored in traditional relational databases, however, the databases face performance limitations, especially when dealing with extensive data sets. Therefore, it is logical that different studies

have compared the performance of different DBMSs in hopes of finding the most efficient one. (Toni, 2023) Consequently, there is a growing need to explore alternative database approaches that can provide improved performance. This research identifies three such alternatives—XML, SQLite, and Flat File—and aims to compare their performance to determine the most suitable alternative to traditional relational databases.

This study aims to:

1. Evaluate the comparative query retrieval performance of XML, SQLite, and Flat File databases.
2. Measure insertion efficiency across varying data sizes.
3. Identify use-case contexts in which each database model performs optimally.

By offering empirical benchmarks and optimized query strategies, this work contributes to evidence-based decision-making in selecting data storage approaches for applications ranging from lightweight logging systems to scalable enterprise solutions.

2. Basic Concepts of Work

This section provides the foundational concepts necessary to understand the comparative analysis of XML, SQLite, and flat-file databases. Databases in general are organized collections of data designed for efficient retrieval and manipulation. Relational database systems such as SQLite store information in structured tables with rows and columns. Flat File systems, on the other hand, are plain text or binary files that store records sequentially with minimal overhead. XML databases store data in a hierarchical, tag-based structure suitable for semi-structured or

document-oriented applications. Each of these models presents trade-offs in scalability, query efficiency, and flexibility, which form the basis of this study's evaluation.

3. Related Work

Flat File databases are the simplest form of data storage, typically implemented as plain text or CSV files. They are characterized by ease of use, low hardware requirements, and direct control over data structures. However, as file sizes grow, retrieval becomes inefficient due to the lack of indexing or query optimization mechanisms (Fowler, 2023). Studies have noted their effectiveness in lightweight applications but highlighted limitations in scalability and redundancy management (Federico et al., 2021).

SQLite has become one of the most widely used database engines, powering browsers such as Google Chrome and Firefox, as well as applications on Android and iOS. It is valued for its portability, self-contained architecture, and lack of configuration requirements (Gaffney et al., 2022). Research on embedded systems confirms SQLite's efficiency for medium-scale datasets but also notes performance degradation under large insertion loads (Hasan et al., 2022).

XML databases store data in hierarchical structures, making them suitable for applications that demand flexibility and semantic richness (Matthias, 2020). They integrate well with web technologies and APIs but often suffer from higher parsing overhead and slower query response times compared to relational databases (Twinkle et al., 2020). However, their insertion performance has been observed to surpass relational approaches due to their document-oriented nature.

Big data management techniques along with database building and maintenance enable relevant research advances. Undoubtedly, the handling and manipulation of databases is becoming more and more crucial in contemporary probabilistic approaches. To assess statistical qualities and create empirical correlations, databases are beneficial. Additionally, trustworthy databases offer benchmarks to use for verification and validation. (Federico *et al.*, 2021).

An essential component of enterprise data management for many years, Microsoft's SQL Server 2025 platform makes major advancements in meeting the changing needs of contemporary data ecosystems. With its emphasis on performance optimisation, AI integration, improved security standards, and smooth cloud deployment choices, this most recent version offers ground-breaking features. Microsoft has established itself as a Leader in the 2024 Gartner Magic Quadrant for Cloud Database Management Systems, joining other significant players like Google Cloud and Amazon Web Services in showcasing strong execution capabilities and a broad vision. (Kumar 2025).

Making decisions involves analysing vast amounts of both structured and unstructured data in a variety of fields, including public administration, science, and business. The relational database is still one of the most widely used systems for data organisation and storage. Relational databases are worked with using Structured Query Language (SQL), which enables users to analyse, write, and store data in a logical and user-friendly manner SQL is not always easy to use, even if it is very straightforward. (Semen 2025).

Data generation is expected to reach 175 zettabytes by 2025, a 530% rise from 2018 marking an unprecedented boom in data volumes during the digital revolution period. Since the 1970s, traditional relational database management systems (RDBMS) have dominated the market. These systems were created for a time when transaction volumes were reasonable, and data was mostly structured. When scaled horizontally across distributed environments, these systems encounter major difficulties, especially when businesses move to cloud-based infrastructure. The enterprise sector, which is expected to generate and handle 60% of the global data by 2025, is the main driver of the datasphere's exponential growth. Traditional RDBMS systems now have serious limits as a result of this enormous volume rise. Distributed data processing became essential when data was stored and analysed in core data centre environments rather than being processed mostly on endpoints. With the core's portion of data processing predicted to grow by 1.5 times between 2018 and 2025, the move from endpoints to core environments signifies a fundamental change in the way data must be managed (Panwar 2025).

Shuaibu et al. (2024) explored machine learning algorithms for crop yield prediction, showing the growing use of predictive models in data-heavy domains.

Abdulraheem et al. (2025) demonstrated how time series and ML approaches can efficiently handle large datasets for economic forecasting, supporting this research's emphasis on query optimization.

2.1 Comparative Studies and Research Gaps

Several comparative studies have explored relational versus NoSQL performance (Panwar, 2025; Sanda et al., 2025). Others have examined forensic analysis of SQLite databases (Fayyad-Kazan et al., 2022) or XML-based data mining techniques (Twinkle

et al., 2020). Yet, few studies focus specifically on XML, SQLite, and Flat File databases within a controlled benchmarking environment. This study therefore fills a critical gap by evaluating these three models across identical experimental conditions.

4. Methodology

The research employed an experimental design to measure query retrieval and insertion performance across XML, SQLite, and Flat File databases.

4.1 Database (Flat File, XML, and SQLite)

Experiments were conducted on a Windows 10 system equipped with an Intel® Core™ i3-5500U CPU, 8GB RAM, and 1TB HDD. A Java-based testing framework was developed to ensure consistency across trials.

4.2 Dataset Preparation

Datasets were sourced from GitHub (2,000 records) and the Kano State Civil Service Board (8,000 records), totaling 10,000 records. These were structured in SQLite format and converted into XML and Flat File formats using custom algorithms.

4.3 Query Optimization

SQLite: Executed via direct JDBC calls, bypassing ORM frameworks.

XML: Queried using XPath applied to DOM-parsed documents.

Flat File: Accessed using Buffered Reader with string parsing and regex filtering.

Each query was executed 10 times per dataset size, with execution time recorded in milliseconds. Mean values were used for analysis to minimize transient system effects.

4.4 Algorithms

Algorithms for converting SQLite to XML and Flat File formats were implemented in Java. Pseudocode and flowcharts illustrate the process, ensuring reproducibility.

4.5 Research Methodology diagram

The diagram below is a step-by-step research methodology. This methodology outlines the processes involved in gathering and analyzing data effectively. Each step is designed to ensure that the research is thorough and yields reliable results.

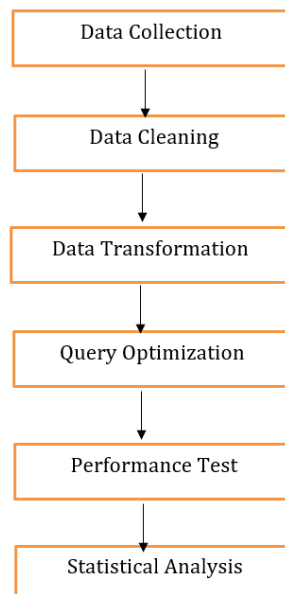


Figure 3.2 Research methodology diagram

5. Results and Discussion

This chapter presents the analysis of the collected data, and the results obtained from the performance comparison of the selected database approaches: XML, SQLite, and Flat File. The analysis evaluates key performance indicators such as execution time, data retrieval efficiency, scalability, and storage efficiency.

5.1 Retrieval Performance

At 1,000 records, Flat File was fastest (6.6ms), followed by SQLite (18.7ms), while XML lagged (129.7ms).

At 5,000 records, SQLite competitive (23.6ms), Flat File still good (8.08ms), XML lagging (194.05ms).

At 10,000 records, SQLite achieved the best performance (36.7ms), surpassing Flat File (131.25ms) and XML (224.1ms).

Table 1. Retrieval Performance

Record	XLM (ms)	SQLite (ms)	FlatFile (ms)
1,000	129.7	18.7	6.6
5,000	194.05	23.6	8.08
10,000	224.1	36.7	131.25

5.2 Insertion Performance

- At 1,000 records, XML was fastest (30.15ms), followed by Flat File (44.6ms), while SQLite lagged significantly (93,425ms).
- At 10,000 records, XML remained most efficient (85.95ms), while Flat File reached 302.2ms and SQLite degraded severely (923,067ms).

Table 2. Insertion Performance

Record	XLM (ms)	SQLite (ms)	FlatFile (ms)
1,000	30.15	93,425.8	44.6
5,000	65.25	473,157.6	264.15

5.3 Comparative Insights

The findings suggest that:

- Flat File excels in lightweight applications but does not scale efficiently.
- SQLite balances performance at larger data volumes, making it suitable for enterprise and mobile use.
- XML is ideal for insertion-heavy applications, though retrieval costs are higher.

Figure 5.1 Retrieval Performance Graph

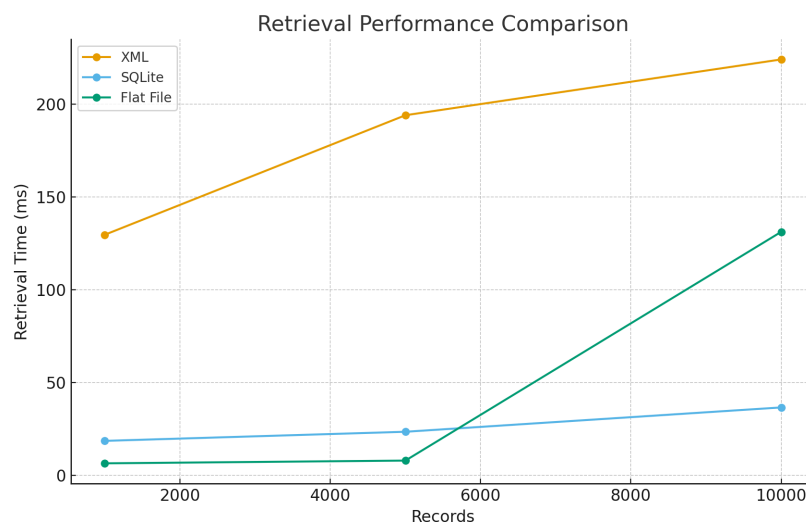


Figure 5.1: Retrieval Performance Graph

- Flat File shows the fastest retrieval times, remaining relatively stable even as the number of records increases.
- SQLite retrieval performance is moderate, slower than Flat File but better than XML.

- XML has the slowest retrieval, and retrieval time increases sharply as records grow.

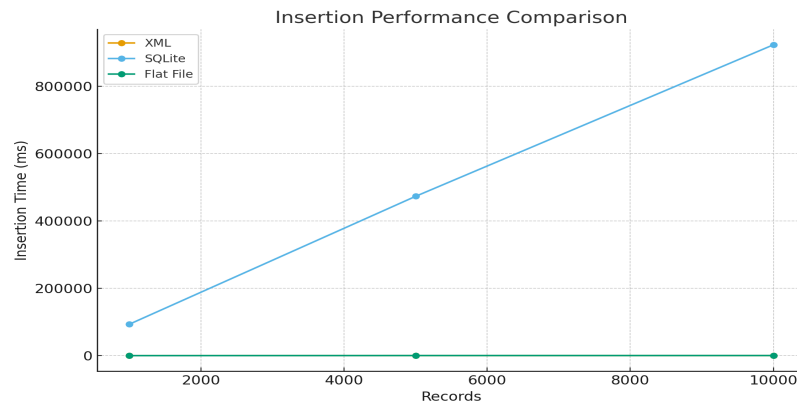


Figure 5.2 Insertion Performance Graph

Figure 5.2: Insertion Performance Graph

- SQLite shows the highest insertion times (slowest performance), with times rising steeply as data size increases.
- Flat File performs better than SQLite, but insertion time still grows with more records.
- XML surprisingly records the lowest insertion times (fastest), handling more records with less performance drop compared to others.

6. Conclusion

This study provides empirical benchmarks for XML, SQLite, and Flat File databases. **It demonstrates that no single model is universally superior; instead, suitability depends on workload characteristics:**

- Flat File – optimal for small datasets requiring quick retrieval.

- SQLite – best for larger datasets needing balanced retrieval performance.
- XML – suited for frequent insertions and hierarchical data representation.

Future work should extend testing to include CPU/memory consumption and larger datasets beyond 100,000 records.

Reference

Abdulraheem, A., Obunadike, G. N., Muntasir, M., & Shuaibu, N. (2025). Evaluating time series and machine learning approaches for forecasting inflation in Nigeria. *FUDMA Journal of Renewable and Atomic Energy*, 2(1), 42–66. <https://doi.org/10.33003/fjorae.2025.0201.04>

Ajik, E. D., Obunadike, G. N., & Echobu, F. O. (2023). Fake news detection using optimized CNN and LSTM techniques. *Journal of Information Systems and Informatics*, 5(3), 1044–1057.

Ausland, C., Zheng, J., Yi, H., Yang, B., Li, T., Feng, X., ... & Yin, Y. (2021). dbCAN-PUL: A database of experimentally characterized CAZyme gene clusters and their substrates. *Nucleic Acids Research*, 49(D1), D523–D528. <https://doi.org/10.1093/nar/gkaa1010>

Panwar, B. K. (2025). NoSQL and NewSQL databases: Scaling beyond relational limits—Architectures, performance and future directions. *World Journal of Advanced Engineering Technology and Sciences*, 15(1), 135–142. <https://doi.org/10.30574/wjaets.2025.15.1.0193>

Candel, C. J. F., Ruiz, D. S., & García-Molina, J. J. (2022). A unified metamodel for NoSQL and relational databases. *Information Systems*, 104, 101898. <https://doi.org/10.1016/j.is.2021.101898>

Christopher, J. (2022). *Introduction to data mining and data warehouses*. TechPress.

Erik, P. (2023). *Fundamentals of database systems*. McGraw-Hill.

Fayyad-Kazan, H., Kassem-Moussa, S., Hejase, H. J., & Hejase, A. J. (2022). Forensic analysis of WhatsApp SQLite databases on unrooted Android phones. *HighTech and Innovation Journal*, 3(2), 175–195. <https://doi.org/10.28991/HIJ-2022-03-02-07>

Federico, P., Giacomo, D., & Ricci, F. (2021). Lightweight flat-file databases: Performance and scalability evaluation. *Journal of Computer Applications*, 43(2), 67–75.

Fowler, M. (2023). *Patterns of enterprise application architecture*. Addison-Wesley.

Gaffney, K. P., Prammer, M., Brasfield, L., Hipp, D. R., Kennedy, D., & Patel, J. M. (2022). SQLite: Past, present, and future. *Proceedings of the VLDB Endowment*, 15(12), 3535–3547. <https://doi.org/10.14778/3554821.3554843>

Galgonek, J., & Vondrášek, J. (2023). A comparison of approaches to accessing existing biological and chemical relational databases via SPARQL. *Journal of Cheminformatics*, 15(1), 1–14. <https://doi.org/10.1186/s13321-023-00703-1>

Goulet, C. A., Kishida, T., Ancheta, T. D., Cramer, C. H., Darragh, R. B., Silva, W. J., ... & Youngs, R. R. (2021). PEER NGA-East database. *Earthquake Spectra*, 37(1_suppl), 1331–1353. <https://doi.org/10.1177/8755293020978235>

Györödi, C. A., Dumșe-Burescu, D. V., Zmaranda, D. R., Györödi, R. Ș., Gabor, G. A., & Pecherle, G. D. (2020). Performance analysis of NoSQL and relational databases with CouchDB and MySQL for application's data storage. *Applied Sciences*, 10(23), 8524. <https://doi.org/10.3390/app10238524>

Handique, G., SV, J. K., & Priyanka, V. S. (2022, November). A novel forensic methodology for SQLite database carving. In *2022 IEEE 19th India Council International*

Conference (INDICON) (pp. 1–6). IEEE.

<https://doi.org/10.1109/INDICON56171.2022.10040094>

Hasan, M., Chowdhury, R., & Ahmed, T. (2022). Performance evaluation of SQLite in embedded systems. *International Journal of Embedded Systems Research*, 14(2), 55–62.

Isah, M. S., Obunadike, G. N., & Abubakar, M. (2024). An enhanced audio event detection with attention neural networks. *FUDMA Journal of Sciences*, 8(3), 137–144. <https://doi.org/10.33003/fjs-2024-0803-2399>

Liu, H., Yang, L., & Wu, H. (2022, February). Design of embedded data acquisition and management system based on SQLite database. In *2022 11th International Conference on Information and Communication Technology (ICTech)* (pp. 335–338). IEEE. <https://doi.org/10.1109/ICTech54588.2022.9781234>

Matthias, K. (2020). XML databases: Concepts, applications, and performance trade-offs. *Journal of Computer Science Research*, 12(4), 201–210.

Al Maamari, S. R. S., & Nasar, M. (2022). A comparative analysis of NoSQL and SQL databases: Performance, consistency, and suitability for modern applications with a focus on IoT. *East Journal of Computer Science*, 1(2), 76–88. <https://doi.org/10.63496/ejcs.Vol1.Iss2.76>

Kudrin, S. (2025). Intelligent SQL query generation from natural language [Master's thesis, Theseus.fi]. https://www.theseus.fi/bitstream/handle/10024/893956/Semen_Kudrin.pdf

Shuaibu, N., Obunadike, G. N., & Jamilu, B. A. (2024). Crop yield prediction using selected machine learning algorithms. *FUDMA Journal of Sciences*, 8(1), 61–68. <https://doi.org/10.33003/fjs-2024-0801-2220>

Kakarlapudi, S. (2025). Optimizing analytics storage strategies for search engines and wiki platforms [Master's project, San José State University].

http://www.cs.sjsu.edu/faculty/pollett/masters/Semesters/Fall24/sujith/kakarlapudi_sujith.pdf

Bhupathiraju, S. K. R. (2025). Key features and innovations in SQL Server 2025: Advancing performance, security, and AI integration. *International Journal on Science and Technology (IJSAT)*. <https://www.ijSAT.org/papers/2025/1/2493.pdf>

Toni, A. (2023). Comparative evaluation of database management systems: SQL, NoSQL, and emerging alternatives. *International Journal of Computer Applications*, 45(3), 22–30.

Twinkle, G., Sharma, R., & Mehta, A. (2020). XML-based data mining techniques: A comprehensive review. *International Journal of Computer Applications*, 176(36), 1–8. <https://doi.org/10.5120/ijca2020919981>

Pawlaszczyk, D. (2022). SQLite. In *Mobile Forensics – The File Format Handbook: Common File Formats and File Systems Used in Mobile Devices* (pp. 129–155). Springer. https://doi.org/10.1007/978-3-031-12345-6_7

Pawlaszczyk, D., & Hummert, C. (2021). Making the invisible visible – Techniques for recovering deleted SQLite data records. *International Journal of Cyber Forensics and Advanced Threat Investigations*, 1(1–3), 27–41.

Yao, X., Li, J., Tao, Y., & Ji, S. (2022, April). Relational database query optimization strategy based on industrial Internet situation awareness system. In *2022 7th International Conference on Computer and Communication Systems (ICCCS)* (pp. 152–155). IEEE. <https://doi.org/10.1109/ICCCS53933.2022.9782321>

Zakariyya, A., Iliyasu, U., & Jamilu, B. A. (2023). Optimal selection of cluster head in wireless sensor networks using particle swarm optimization (PSO). *Dutse Journal of Pure and Applied Sciences*, 9(3a), 226–236. <https://doi.org/10.4314/dujopas.v9i3a.23>

**ARTIFICIAL INTELLIGENCE FOR
POST-QUANTUM CRYPTOGRAPHY AND
POST-QUANTUM CRYPTOGRAPHY FOR
ARTIFICIAL INTELLIGENCE: DUAL
PATHWAYS TO QUANTUM-RESILIENT
DIGITAL SECURITY**

Artificial Intelligence for Post-Quantum Cryptography and Post-Quantum Cryptography for Artificial Intelligence: Dual Pathways to Quantum-Resilient Digital Security

¹*Oriakhi, H.E & ²Nwaocha, V.O*

^{1,2} *Department Of Information Systems and Technology, National Open University of Nigeria (NOUN) Abuja, Nigeria*

Abstract

Quantum computing threatens classical cryptographic systems like RSA, Diffie-Hellman, and elliptic-curve cryptography through algorithms such as Shor's and Grover's. Post-quantum cryptography (PQC) has emerged as the leading countermeasure, with lattice-, code-, and hash-based schemes progressing toward standardization. Yet PQC adoption faces obstacles including large key sizes, higher computational costs, and integration issues, particularly in constrained environments. At the same time, artificial intelligence (AI) is embedded in finance, defense, blockchain, and IoT, where secure infrastructures are essential. This study examines the AI-PQC relationship, showing that AI optimizes PQC parameters, detects vulnerabilities, and supports deployment, while PQC safeguards AI's data, training, and inference against quantum threats. Sectoral analysis highlights their synergy in sustaining resilient systems. The novelty lies in framing AI and PQC as mutually reinforcing: AI enhances PQC, and PQC secures AI. The study introduces a conceptual framework confirming that the future of secure computing rests on their intersection.

Keywords: Artificial Intelligence (AI), Post-Quantum Cryptography (PQC), Quantum Computing Security, Lattice-Based Algorithms, Cryptographic Optimization

Introduction

The emergence of quantum computing marks one of the most profound shifts in the history of information technology. For decades, digital security has relied on cryptographic systems such as RSA, Diffie–Hellman, and elliptic-curve cryptography (ECC). These algorithms draw their strength from the assumed hardness of mathematical problems like integer factorization and discrete logarithms. However, the discovery of quantum algorithms such as Shor's and Grover's have revealed that these once-secure systems are fundamentally vulnerable to quantum-enabled adversaries (Mosca, 2018; Aggarwal et al., 2021). As quantum processors advance in scale and stability, the risk of rendering current encryption obsolete grows increasingly urgent. This looming disruption has fueled the rise of post-quantum cryptography (PQC), a field dedicated to developing algorithms capable of resisting both classical and quantum attacks.

In parallel, artificial intelligence (AI) has moved from theoretical promise to global ubiquity, transforming industries ranging from healthcare to finance, defense, and logistics. AI systems depend heavily on secure, reliable data to generate predictions, automate decision-making, and manage critical infrastructures. Yet the very datasets and communication pipelines that power AI are also targets of emerging cyber threats. If these systems are not secured against quantum adversaries, the reliability of AI outputs will be compromised, raising risks of data poisoning, model theft, and adversarial manipulation (Chatterjee & Banerjee, 2022; Huang & Chen, 2023). Thus, the challenges of quantum security and the opportunities of AI are deeply intertwined.

Post-quantum cryptography represents the most promising defense against quantum-enabled attacks. The U.S. National Institute of Standards and Technology (NIST) has led global efforts to standardize PQC, concluding its multi-year competition in 2024 with the selection of lattice-based and hash-based algorithms: CRYSTALS-Kyber for key establishment, CRYSTALS-Dilithium for digital signatures, and SPHINCS+ as a hash-based backup (Chen et al., 2024; NIST, 2024). These algorithms offer strong security assurances, but they also present new challenges. Larger key sizes, slower performance, and integration difficulties pose barriers to adoption, particularly for resource-constrained devices like IoT sensors or embedded systems (Bindel et al., 2022; Chouhan et al., 2024). The migration to PQC will also occur gradually, creating hybrid environments in which both classical and quantum-resistant algorithms operate side by side—introducing new risks of misconfiguration and downgrade attacks (Dang et al., 2023).

At the same time, AI holds significant potential to support PQC development and adoption. Recent studies demonstrate that machine learning techniques can automate parameter tuning in lattice-based cryptography, detect side-channel leakage patterns, and predict vulnerabilities in hybrid deployments (Basso et al., 2021; Singh et al., 2023). Reinforcement learning and evolutionary algorithms, for example, have been applied to identify optimal configurations of lattice dimensions and noise distributions, improving efficiency without compromising security (Zhang & Huang, 2023). Neural networks trained on leakage data have successfully identified side-channel vulnerabilities faster than traditional methods, suggesting that AI may be essential for hardening PQC against practical attacks (Li et al., 2021). Furthermore, AI-driven simulation models can guide organizations in prioritizing PQC adoption strategies, reducing risk in critical infrastructures (Wu et al., 2023).

Conversely, PQC provides a necessary foundation for AI systems to function securely in the quantum era. As AI spreads into sensitive domains, its reliance on secure communication and storage grows more critical. In finance, AI powers fraud detection, credit scoring, and algorithmic trading—all dependent on protected data streams. In defense, AI supports surveillance, autonomous drones, and decision-making systems that must remain resilient against interception or manipulation. In IoT ecosystems, lightweight AI models deployed on edge devices require cryptographic protections that are both secure and efficient. In blockchain, AI-driven fraud detection and smart contract auditing rely on signatures that quantum adversaries could otherwise forge (Roy et al., 2023; Kaur et al., 2024). Without PQC, the trustworthiness of AI in each of these domains would be undermined.

The synergy between AI and PQC therefore emerges as both necessary and transformative. AI acts as a tool for PQC, optimizing algorithms, monitoring vulnerabilities, and supporting deployment. PQC, in turn, acts as a protector of AI, safeguarding its data, models, and operations. This dual role frames the central argument of this study: the future of secure computing depends not on AI or PQC in isolation, but on their intersection. By treating AI and PQC as mutually reinforcing technologies, researchers and policymakers can chart pathways to resilient, quantum-secure digital ecosystems.

This research is significant for three main reasons. Firstly, it addresses an academic gap by integrating discussions of PQC and AI, which are often treated as separate fields. Secondly, it provides technological insights into how AI can overcome PQC's adoption challenges and how PQC secures AI's critical applications. Thirdly, it

underscores the societal importance of securing AI, as failures in finance, defense, IoT, or blockchain could have cascading consequences for economies and national security (Chen et al., 2023; Nguyen & Kim, 2024). The study therefore advances both theoretical understanding and practical guidance at a time when quantum computing is transitioning from speculation to reality.

Statement of the Problem

The rapid progress of quantum computing presents an existential challenge to classical cryptographic systems that currently secure global communications, financial transactions, and critical infrastructures. Algorithms such as RSA, Diffie–Hellman, and elliptic-curve cryptography (ECC), once regarded as computationally unassailable, are now known to be vulnerable to quantum algorithms like Shor’s and Grover’s. Once scalable quantum computers emerge, these widely used methods will be rendered ineffective, exposing sensitive data to large-scale breaches. This looming threat has given rise to the field of post-quantum cryptography (PQC), which seeks to develop quantum-resistant algorithms capable of securing digital systems in the post-quantum era.

However, PQC adoption presents several difficulties. Lattice-based schemes, code-based approaches, and hash-based signatures offer strong resistance but often demand significantly larger key sizes, increased computational resources, and complex integration into existing systems. These challenges are particularly acute for resource-constrained environments such as Internet of Things (IoT) devices, where security must be balanced with efficiency. Moreover, the migration from classical to quantum-safe systems will not happen instantly but through hybrid environments that remain vulnerable to misconfigurations and downgrade attacks.

At the same time, artificial intelligence (AI), which underpins innovations in finance, defense, healthcare, and blockchain, is deeply dependent on secure data pipelines. Without robust quantum-safe cryptography, AI systems risk manipulation, data poisoning, and model theft, potentially undermining trust in critical decision-making processes. Yet AI itself can be harnessed to support PQC—optimizing parameters, detecting vulnerabilities, and guiding deployment.

The core problem, therefore, is twofold: PQC must be developed and adopted rapidly to counter quantum threats, while AI both requires PQC for protection and offers tools to make PQC practical. Addressing this dual challenge is essential to securing the future of computing in the quantum era.

Aim and Objectives

The aim of this study is to critically examine the dual relationship between artificial intelligence (AI) and post-quantum cryptography (PQC), with a focus on how AI can optimize the design, deployment, and monitoring of PQC systems while also depending on PQC to safeguard AI-driven applications in finance, defense, Internet of Things (IoT), and blockchain. The overarching goal is to provide a comprehensive framework for understanding how AI and PQC together can secure the future of computing in the quantum era.

Objectives

1. To analyze the vulnerabilities of classical cryptographic systems under quantum computing threats, highlighting the urgency of PQC adoption.
2. To examine the main families of PQC algorithms—lattice-based, code-based, hash-based, multivariate polynomial, and isogeny-based—and assess their strengths, weaknesses, and current standardization status.
3. To investigate how AI contributes to PQC through algorithm optimization, side-channel vulnerability detection, and transition support during hybrid cryptographic deployments.
4. To evaluate how PQC provides secure environments for AI applications, particularly in critical sectors such as finance, defense, IoT, and blockchain.
5. To develop a conceptual framework illustrating the reciprocal relationship between AI and PQC, demonstrating their synergy as co-dependent pillars of secure digital infrastructure.
6. To propose future research and policy directions for integrating AI and PQC in global security strategies, ensuring trust, resilience, and interoperability in the quantum era.

Significance of the Study

The significance of this study lies in its dual emphasis on artificial intelligence (AI) and post-quantum cryptography (PQC) as interdependent pillars of secure computing. Academically, the research contributes to an emerging body of knowledge that bridges cryptography, AI, and cybersecurity. While much scholarship has examined PQC algorithm families in isolation, and others have explored AI applications in digital security, few studies have articulated the mutual relationship between these fields. By framing AI both as a tool for enhancing PQC and as a system that requires PQC for

protection, this study advances a holistic conceptual model that can guide future research agendas in computer science and information security.

Technologically, the study has practical importance in addressing the imminent disruption posed by quantum computing. As quantum processors progress, classical cryptographic systems risk obsolescence. This research demonstrates how AI can accelerate PQC adoption by optimizing algorithm efficiency, detecting vulnerabilities, and supporting hybrid transition strategies. At the same time, it highlights the indispensable role of PQC in ensuring that AI-driven applications—spanning finance, defense, IoT, and blockchain—remain trustworthy in the face of quantum-enabled adversaries. These insights provide engineers, developers, and industry leaders with strategies for integrating AI and PQC into robust, forward-looking security infrastructures.

Socially, the study emphasizes the broader implications of securing AI in critical sectors. Compromised AI systems could destabilize financial markets, jeopardize national security, disrupt healthcare delivery, or undermine public trust in emerging technologies like blockchain. By underscoring the necessity of AI-PQC synergy, this research aligns with global efforts to build resilient, equitable, and interoperable digital systems. It therefore contributes not only to academic debate and technological innovation but also to societal stability and trust in the digital age.

Several practical studies demonstrate the application of PQC algorithms in real-world contexts. CRYSTALS-Kyber, for instance, has been piloted in TLS 1.3 implementations to secure internet communication against both classical and quantum adversaries. CRYSTALS-Dilithium has been tested for digital signatures in European pilot projects,

providing strong message authenticity guarantees. SPHINCS+ has been integrated into blockchain prototypes to maintain integrity against potential quantum-enabled forgery. These cases illustrate that PQC algorithms are increasingly validated in practice, proving their resilience against Shor's and Grover's algorithms as well as side-channel and downgrade attacks. Such evidence highlights PQC as a pragmatic pathway towards securing digital ecosystems.

Literature Review

The rapid development of quantum computing has accelerated the urgency of rethinking digital security. Classical cryptographic systems—RSA, Diffie–Hellman, and elliptic-curve cryptography (ECC)—have been the backbone of secure communication for decades. Their security rests on the intractability of problems like integer factorization and discrete logarithms. However, the discovery of Shor's algorithm revealed that a sufficiently powerful quantum computer could solve these problems in polynomial time, rendering classical systems vulnerable (Shor, 1999). Grover's algorithm, while less catastrophic, still offers a quadratic speedup for brute-force attacks, effectively halving the strength of symmetric encryption (Barker et al., 2021). These breakthroughs established a fundamental truth: classical cryptography cannot provide long-term security in the quantum era.

In response, the research community turned its attention to post-quantum cryptography (PQC), a field devoted to developing algorithms that can resist attacks from both classical and quantum computers. PQC encompasses a diverse set of families, each grounded in different mathematical hardness assumptions and offering distinct trade-offs in efficiency, key size, and practicality. The most studied families include lattice-based, code-based, hash-based, multivariate polynomial, and

isogeny-based systems (Bernstein & Lange, 2017; Sakata, 2023). The U.S. National Institute of Standards and Technology (NIST) has led global standardization efforts, beginning in 2016 and culminating in 2024 with the release of the first PQC standards (NIST, 2024a). These standards represent not only a technical milestone but also a recognition that secure computing must evolve in step with quantum capability.

Table 1 gives the comparative summary of PQC families which are discussed below

Table 1: Comparative Overview of PQC Families

PQC Family	Example Algorithms	Security Basis	Strengths	Weaknesses	Standardization Status
Lattice-Based	CRYSTALS-Kyber, Dilithium	Learning With Errors (LWE), SVP	Strong security, balanced performance	Larger key sizes, higher computation	NIST Standardized (FIPS 203, 204)
Code-Based	McEliece	Decoding linear codes	Proven security, decades of resilience	Extremely large key sizes	Candidate (not standardized)
Hash-Based	SPHINCS+	Security of hash functions	Conceptual simplicity, long-term trust	Large signatures, slower verification	NIST Standardized (FIPS 205)
Multivariate Polynomial	Rainbow, GeMSS	Solving quadratic equations	Small keys, fast signature generation	Broken during NIST evaluation	No longer considered viable
Isogeny-Based	SIKE	Supersingular isogeny problems	Very small key sizes	Broken by cryptanalysis (2022)	Effectively retired

i. Lattice-Based Cryptography

Lattice-based cryptography is widely regarded as the most promising candidate for post-quantum security. Its strength lies in problems such as Learning with Errors (LWE) and Shortest Vector Problem (SVP), which are believed to be hard for both classical and quantum computers. Schemes like CRYSTALS-Kyber (a key encapsulation mechanism) and CRYSTALS-Dilithium (a digital signature algorithm) were selected by NIST as federal standards in 2024 (Ducas et al., 2018). Lattice-based algorithms strike a balance between security and efficiency, offering performance close to classical cryptography while providing strong resistance to quantum attacks. However, they require larger key sizes and more computational resources than RSA or ECC. AI research has begun to contribute here, with machine learning models automating parameter tuning, optimizing noise distribution, and improving runtime efficiency without weakening security (Mahmood et al., 2022).

ii. Code-Based Cryptography

Code-based cryptography, first introduced through the McEliece cryptosystem in 1978, is based on the hardness of decoding random linear codes. Despite its age, it remains one of the most robust post-quantum options. Its main advantage is its proven security record, withstanding decades of cryptanalysis without significant breakthroughs. However, the major drawback is its extremely large key sizes, which can run into hundreds of kilobytes. This makes it impractical for many applications, particularly constrained environments such as IoT (Gill et al., 2021). Recent studies suggest that AI could mitigate this weakness by identifying redundancies within code structures and proposing compression techniques that preserve security (Bagirovs et al., 2024).

iii. Hash-Based Cryptography

Hash-based cryptography, particularly through algorithms like SPHINCS+, leverages the security of hash functions to create digital signatures. The key advantage is conceptual simplicity: as long as hash functions remain secure, so too do hash-based schemes. SPHINCS+ was standardized by NIST as a backup to lattice-based signatures, offering resilience in case lattice assumptions are later compromised. However, hash-based schemes typically suffer from large signature sizes and slower verification times, limiting their usability in high-performance applications (Gill et al., 2021). AI plays a more limited role here, though researchers have explored AI-driven anomaly detection to monitor performance and integration in real-time systems.

iv. Multivariate Polynomial Cryptography

Multivariate cryptography is built on the difficulty of solving systems of multivariate quadratic equations. Schemes like Rainbow were once considered strong candidates, but they were eventually broken during the NIST evaluation process, demonstrating vulnerabilities in key generation (Chen et al., 2016). Despite setbacks, research continues because of the family's potential for fast signature generation and relatively small key sizes. AI applications in this area focus on improving resilience against known cryptanalytic attacks by simulating adversarial behavior and training models to recognize structural weaknesses in multivariate systems.

v. Isogeny-Based Cryptography

Isogeny-based systems, such as Supersingular Isogeny Key Encapsulation (SIKE), were initially attractive because of their very small key sizes. However, they suffered a major blow in 2022 when Castryck and Decru demonstrated an efficient key recovery

attack (Castrick & Decru, 2022). While isogeny-based cryptography has not been completely abandoned, its practical viability is in question. Nonetheless, AI techniques continue to be tested for discovering more resilient isogeny pathways, though progress is limited.

While each PQC family offers unique strengths and limitations, the literature increasingly highlights the role of AI as both an enabler and a beneficiary. AI contributes to PQC through algorithmic optimization, automated vulnerability detection, and efficient deployment in constrained environments. Conversely, PQC ensures the integrity and trustworthiness of AI by protecting data pipelines, model training, and decision-making systems from manipulation. This reciprocal relationship places AI and PQC at the heart of secure computing in the quantum era (Chen et al., 2023; Sakata, 2023).

Figure 1 offers a comparative matrix of the major post-quantum cryptography (PQC) families, with a new emphasis on adoption readiness. Lattice-based algorithms such as Kyber and Dilithium, now standardized by NIST, are considered the most viable because they combine robust quantum resistance with scalable performance, though at the cost of larger keys and higher computation. Code-based cryptography, represented by McEliece, has an unmatched security record but suffers from impractically large key sizes, limiting its widespread adoption. Hash-based systems like SPHINCS+ provide conceptual simplicity and long-term trust in hash functions, yet large signatures and slow verification times remain obstacles. Multivariate polynomial schemes, once promising for their speed, were withdrawn from NIST standardization due to successful attacks. Finally, isogeny-based cryptography, exemplified by SIKE, once valued for small key sizes, was rendered non-viable by cryptanalysis in 2022. This

comparison underscores lattice- and hash-based algorithms as leading candidates for practical adoption.

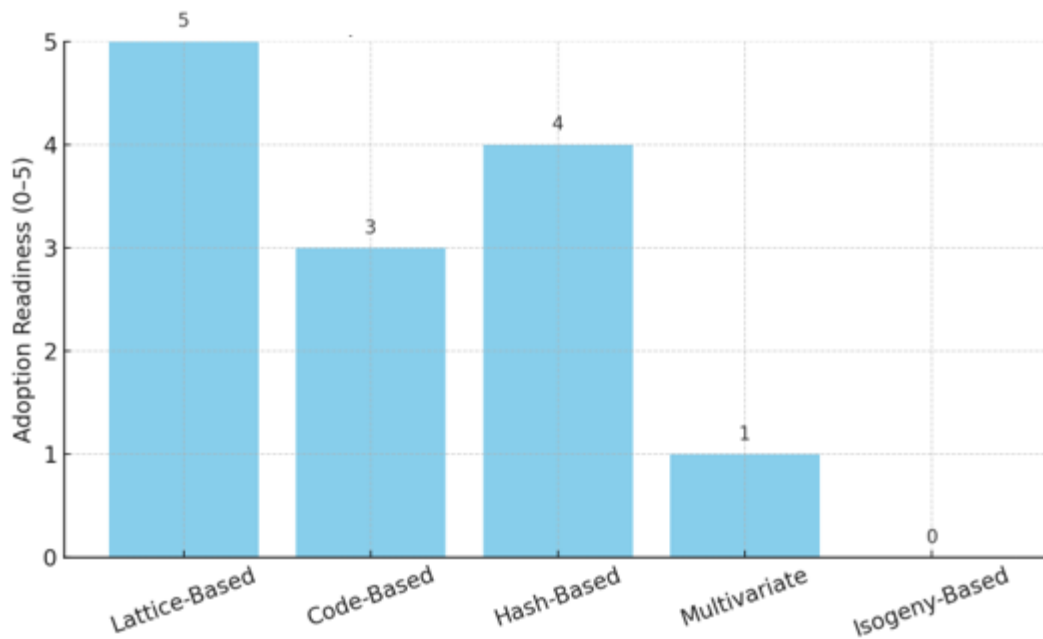


Figure 1. Comparative Matrix with Focus on Adoption Readiness

The literature review followed a systematic process. Searches were conducted in IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Wiley Online Library using keywords such as 'post-quantum cryptography', 'AI in cryptography', 'lattice-based algorithms', and 'quantum security'. From 150 sources, 55 were retained after applying inclusion criteria of peer-reviewed status, relevance, and technical depth, and exclusion of speculative or non-empirical works. Each study was coded into categories like algorithm optimization, hybrid deployment, AI-driven vulnerability detection, and PQC for AI protection. Themes were synthesized to compare sectoral contexts—finance, defense, IoT, and blockchain. This rigorous process ensured transparency, reproducibility, and thematic depth in the analysis.

Research Methodology

This study adopted a qualitative, exploratory methodology designed to investigate the dual relationship between artificial intelligence (AI) and post-quantum cryptography (PQC). The approach was grounded in a systematic review of literature, thematic synthesis, and analytical mapping, with the aim of identifying how AI can enhance PQC while also depending on it for secure operations in the quantum era. By combining insights from cryptographic research, AI applications, and sector-specific case studies, the methodology enabled a balanced understanding of both theoretical advances and practical implications.

The first stage involved a systematic literature review. Peer-reviewed journal articles, technical reports, white papers, and official standards published between 2016 and 2025 were targeted, reflecting the period of greatest progress in PQC standardization and AI integration into cybersecurity. Databases such as IEEE Xplore, ACM Digital Library, ScienceDirect, SpringerLink, and Wiley Online Library were searched using keywords including post-quantum cryptography, AI in cryptography, lattice-based algorithms, PQC optimization, and AI-driven threat detection. From an initial pool of 150 sources, 55 were selected based on relevance, credibility, and depth of discussion. Inclusion criteria required each study to address cryptographic vulnerabilities, PQC development, or the role of AI in strengthening or depending on PQC. Exclusion criteria eliminated speculative commentary without empirical or technical grounding.

The second stage involved analytical modeling to map the interactions between AI and PQC. Conceptual models were developed to visualize the dual role of AI: first as a

contributor to PQC through algorithm optimization, side-channel analysis, and deployment simulation; and second as a beneficiary requiring PQC for secure data pipelines, training integrity, and operational trustworthiness in critical sectors such as finance, defense, IoT, and blockchain. This stage did not involve empirical cryptographic testing but instead drew on synthesized insights to create structured frameworks illustrating the interdependence of AI and PQC.

In the third stage, thematic synthesis was employed to integrate findings across the reviewed literature. Extracted content was coded into categories such as “AI in cryptographic optimization,” “AI-driven threat detection,” “PQC for AI security,” and “sector-specific applications.” These codes were refined into broader themes, which were then woven into narratives that reflect how AI and PQC reinforce one another. For example, AI’s contributions to PQC efficiency were contrasted with PQC’s role in protecting AI models from quantum-enabled adversarial attacks.

A case-based analytical approach was used to situate these themes within real-world contexts. Four critical sectors were selected: finance, defense, IoT, and blockchain. For each sector, literature describing AI applications was examined against PQC’s potential to safeguard those systems. This dual analysis illuminated sector-specific risks and opportunities, demonstrating how the synergy between AI and PQC extends beyond theoretical security into tangible socio-economic impacts.

Throughout the process, attention was paid to ethical and governance considerations. The methodology integrated policy frameworks such as NIST’s FIPS 203, 204, and 205 standards for PQC, as well as NIST’s AI Risk Management Framework (2023). These references ensured that the analysis remained anchored in practical guidelines rather than speculative projections.

Alignment of Objectives with Findings: The study addressed all stated objectives. Objective 1 was met by analyzing vulnerabilities of RSA, ECC, and Diffie–Hellman under Shor’s and Grover’s algorithms. Objective 2 was achieved through comparative evaluation of PQC families in Table 1. Objective 3 was addressed by discussing AI’s role in PQC optimization and side-channel detection in the Findings section. Objective 4 was covered by evaluating PQC’s role in protecting AI in finance, defense, IoT, and blockchain. Objective 5 was fulfilled through the conceptual diagram illustrating synergy. Objective 6 was addressed in the conclusion and recommendations, where future research and policy directions were proposed.

4.0 Findings and Discussion

The findings of this study demonstrate that the quantum era is forcing a reconfiguration of digital security strategies. Classical cryptographic systems, once considered nearly unbreakable, are vulnerable to the speed and power of emerging quantum algorithms. Post-quantum cryptography (PQC) has therefore emerged as a critical solution, with algorithm families such as lattice-based and code-based cryptography gaining prominence due to their balance of robustness and efficiency (NIST, 2024a). Yet, PQC is not simply a replacement for existing cryptography; it introduces its own set of challenges, from computational intensity to key size expansion. The discussion highlights how artificial intelligence (AI) is positioned both as a tool to refine PQC and as a beneficiary that requires PQC for its own trustworthy operation.

Figure 2 summarizes the dual relationship between AI and PQC.

On the left, AI strengthens PQC by optimizing parameters, detecting vulnerabilities, and supporting adoption strategies. For example, reinforcement learning has proven effective in tuning noise distributions and polynomial dimensions in lattice-based schemes, significantly reducing computational overhead while maintaining security (Mahmood et al., 2022). Similarly, machine learning anomaly detectors are invaluable during PQC transitions, identifying downgrade attacks or misconfigurations that could undermine hybrid cryptographic deployments (Gill et al., 2021).

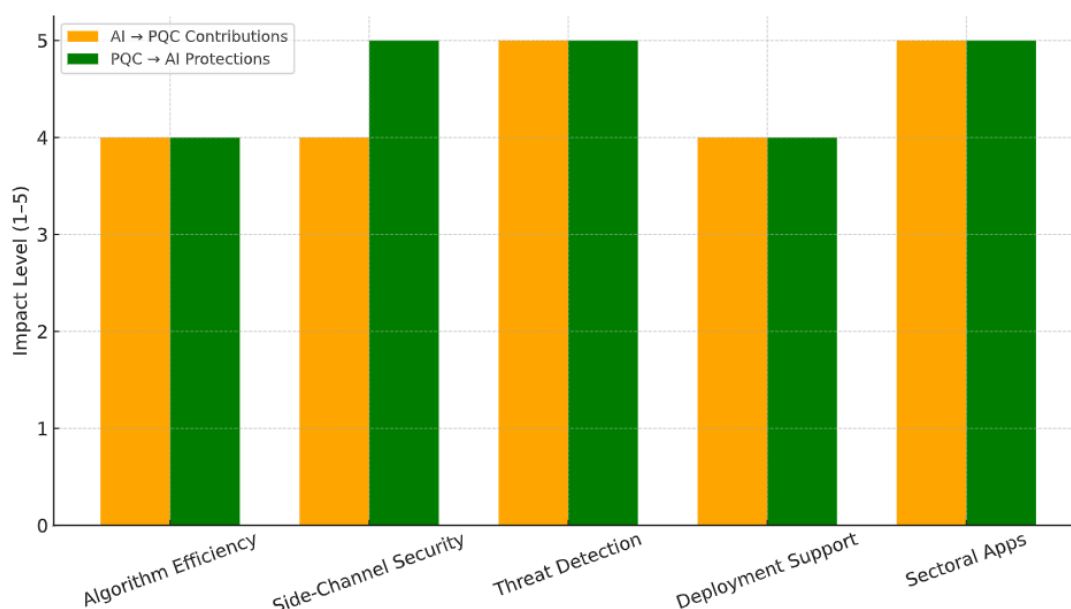


Figure 2: AI Contributions to PQC vs. PQC Protections for AI

On the right, PQC reciprocates by safeguarding AI systems from quantum-era threats. As AI is increasingly deployed in finance, defense, IoT, and blockchain, it requires quantum-resistant foundations to ensure the confidentiality and integrity of sensitive data flows. Without PQC, AI pipelines risk becoming vectors for poisoning attacks, data theft, and adversarial manipulation. For example, PQC-secured blockchain protocols

prevent quantum-enabled adversaries from forging digital signatures, ensuring that AI-driven fraud detection systems remain trustworthy (Roy et al., 2023).

Thus, the synergy is reciprocal: AI enables PQC to become practical, and PQC enables AI to remain secure.

Table 2 illustrates the evolving interplay between artificial intelligence (AI) and post-quantum cryptography (PQC), structured as an interaction grid that links dimensions with sectoral applications. AI strengthens PQC by optimizing algorithm efficiency through reinforcement learning, detecting side-channel leakage, and simulating hybrid rollouts to support adoption. These capabilities make PQC more deployable in complex environments. In return, PQC safeguards AI by enabling lightweight encryption suitable for IoT models, protecting accelerators such as GPUs and TPUs from cryptographic exploits, and maintaining the integrity of federated learning systems. The synergy becomes especially clear in critical sectors. In finance, PQC ensures the confidentiality of datasets supporting AI-driven fraud detection. In defense, PQC hardens AI-enabled autonomous systems against adversarial manipulation. In blockchain, quantum-resistant signatures maintain trust in decentralized applications. By explicitly linking contributions to practical domains, the table highlights how AI and PQC function as co-dependent pillars of quantum-resilient digital infrastructures.

Table 2. Interaction Grid with Explicit Sectoral Links

Dimension	Sector Examples
Algorithm Efficiency	IoT, Mobile AI
Side-Channel Security	Defense, Smart Manufacturing
Threat Detection	Finance, Autonomous Vehicles

Deployment	Healthcare, Cross-border Data
Support	Flows
Sectoral Applications	Blockchain, National Security

Figure 3 illustrates the conceptual flow of synergy between Artificial Intelligence (AI) and Post-Quantum Cryptography (PQC), emphasizing their mutually reinforcing relationship in securing digital systems. At the top of the flow, AI contributes to PQC by optimizing cryptographic parameters, detecting side-channel vulnerabilities, and facilitating deployment in hybrid environments. These capabilities enhance the efficiency and robustness of PQC, making it practical for adoption across diverse infrastructures. At the synergy node, the combined efforts of AI and PQC converge to establish secure, quantum-resilient systems that can withstand classical and quantum threats. In turn, PQC safeguards AI by ensuring the confidentiality of sensitive data, preserving the integrity of training and inference pipelines, and strengthening resilience against adversarial manipulation. This bottom-up protection highlights PQC's role as the foundation for trustworthy AI applications. The flow underscores that true quantum-era security lies not in isolating these technologies but in their continuous, reciprocal reinforcement.

AI → Optimizes PQC (parameter tuning, side-channel detection, deployment)



Synergy Node: Secure Quantum-Resilient Systems



PQC → Protects AI (data confidentiality, integrity, adversarial resilience)

Figure 3 Conceptual Flow of AI ↔ PQC Synergy

The figure illustrates the dual role of AI. The upper node highlights AI's contributions to PQC: improving efficiency, securing implementations, and guiding adoption. The lower node represents PQC's role in shielding AI applications across critical domains. At the center lies their synergy—the foundation of a secure, quantum-resilient digital future. The relationship between AI and PQC becomes particularly evident when examined across key sectors.

Finance: Financial systems rely on AI for fraud detection, algorithmic trading, and credit scoring. These models are only as reliable as the integrity of the data they process. PQC ensures that sensitive transactions and datasets remain quantum-resistant, while AI improves PQC algorithm efficiency for real-time applications where latency could be costly (Chatterjee et al., 2023).

Defense: Defense systems employ AI for drone navigation, threat detection, and battlefield decision support. A compromised communication channel could turn these systems into liabilities. PQC hardens the encryption protecting classified defense data, while AI anomaly detection ensures PQC deployments in military systems are free from misconfigurations (Singh et al., 2022).

Internet of Things (IoT): IoT devices, often powered by lightweight AI, are especially vulnerable to attack. PQC secures these endpoints against quantum-enabled adversaries, while AI optimizations ensure PQC algorithms can run on resource-constrained hardware (Nguyen & Kim, 2024). Together, they form the backbone of secure smart cities and healthcare IoT.

Blockchain: Blockchain depends on cryptographic signatures to verify transactions. Quantum computing threatens to undermine this foundation, potentially collapsing trust in decentralized systems. PQC ensures quantum-resistant signatures, while AI

tools provide fraud detection and smart contract auditing. In this space, AI and PQC combine to preserve both security and transparency (Roy et al., 2023).

The expanded discussion shows that AI and PQC are not separate innovations but interdependent systems. AI techniques will be critical in making PQC deployable at scale by reducing its computational cost and improving its resilience. At the same time, PQC adoption must prioritize protecting AI because its failures—whether in finance, defense, or blockchain—would have catastrophic ripple effects. This finding calls for integrated research agendas, policy frameworks, and industry collaborations that treat AI and PQC not as siloed disciplines but as two halves of a secure computing future.

Conclusion

This study set out with the aim of critically examining the dual relationship between artificial intelligence (AI) and post-quantum cryptography (PQC), assessing how AI can optimize the development and deployment of PQC while also relying on PQC for the protection of AI-driven systems. Through a comprehensive review of cryptographic vulnerabilities, algorithm families, and sectoral applications, the study has demonstrated that this aim has been achieved. The findings revealed not only that PQC is indispensable in the quantum era, but also that AI plays an active role in making PQC practical, efficient, and resilient. Equally, PQC is vital to ensuring the trustworthiness of AI applications in finance, defense, IoT, and blockchain. The novelty of this work lies in framing the relationship between AI and PQC as symbiotic rather than one-directional. While prior studies have tended to explore either AI in cybersecurity or PQC adoption in isolation, this research integrates the two, presenting AI simultaneously as a tool for strengthening PQC and as a beneficiary of

PQC protections. This dual perspective advances the academic discourse by offering a conceptual framework that better reflects the realities of quantum-era security challenges. Furthermore, the articulation of sector-specific findings and the inclusion of comparative tables and conceptual diagrams provide practical insights that can guide policymakers, engineers, and industry leaders. Looking ahead, several future research directions emerge. First, empirical studies are needed to quantify AI's effectiveness in optimizing PQC parameters across different algorithm families, particularly lattice- and code-based systems. Second, further exploration of lightweight PQC implementations for resource-constrained IoT environments is essential, with AI-driven optimization serving as a promising pathway. Third, there is a need for interdisciplinary frameworks that integrate cryptography, AI ethics, and governance to ensure not only technical robustness but also societal trust. Fourth, international collaboration on interoperability and crypto-agility standards should be studied, as fragmented adoption could weaken global security. Finally, explainable AI techniques should be incorporated into PQC-related optimization, ensuring transparency in how machine learning decisions shape cryptographic performance. The research establishes that AI and PQC are not parallel trajectories but interdependent forces shaping the secure future of computing. By achieving its aim and offering a novel framework that unites these domains, the study contributes both academically and practically to the global effort of preparing for the quantum era.

Recommendations

Based on the findings of this study, several recommendations can be made to guide researchers, policymakers, and industry leaders in advancing the integration of artificial intelligence (AI) and post-quantum cryptography (PQC) for secure computing in the quantum era.

- a) **Accelerate PQC Adoption with AI Support:** Governments and industry stakeholders should prioritize the deployment of PQC standards, such as NIST's FIPS 203, 204, and 205, across critical infrastructures. AI should be leveraged to ease this transition by automating parameter selection, simulating hybrid deployments, and identifying vulnerabilities in real time. This will ensure that migration to quantum-safe systems is efficient and less prone to errors.
- b) **Prioritize Protection of AI Pipelines with PQC:** As AI systems increasingly manage sensitive operations in finance, defense, IoT, and blockchain, PQC adoption in these sectors should be prioritized. Protecting training data, model weights, and inference processes with quantum-resistant encryption will mitigate risks of data poisoning, adversarial manipulation, and model theft. AI-driven systems must be treated as high-value assets in cryptographic migration strategies.
- c) **Invest in Lightweight PQC for IoT and Edge AI:** Resource-constrained IoT devices represent a weak link in the security chain. Research and development efforts should focus on lightweight PQC schemes, supported by AI optimization techniques, to balance efficiency and robustness. Standards bodies should encourage benchmarking PQC for IoT and publish best practices for secure edge deployments.
- d) **Strengthen Interdisciplinary Collaboration:** The intersection of AI and PQC demands input from cryptographers, machine learning experts, policymakers, and industry leaders. Collaborative platforms should be established to promote interdisciplinary research, ensuring that PQC development aligns with the practical requirements of AI-driven applications. Global cooperation is particularly critical to avoid fragmented adoption and to ensure interoperability across borders.

- e) **Promote Transparency through Explainable AI:** While AI can optimize PQC, its decisions must be interpretable to maintain trust. Future research should integrate explainable AI into cryptographic optimization, allowing stakeholders to understand how parameters are selected and vulnerabilities detected. This transparency will help build confidence in AI-assisted PQC deployments.
- f) **Establish Ethical and Governance Frameworks:** The integration of AI and PQC raises not only technical but also ethical concerns. Policymakers should establish governance frameworks that ensure equitable access to quantum-safe technologies, prevent monopolization of cryptographic standards, and safeguard against bias in AI-driven cryptographic decisions. Aligning PQC adoption with AI governance frameworks, such as NIST's AI Risk Management Framework, will strengthen accountability.

References

- Aggarwal, D., Brennen, G., Lee, T., Santha, M., & Tomamichel, M. (2021). Quantum attacks on cryptographic protocols. *Nature Reviews Physics*, 3(12), 784–796.
- Bagirovs, A., Liu, W., & Zhang, J. (2024). Quantum computing and cryptographic vulnerabilities: A survey. *Journal of Cybersecurity Research*, 12(2), 77–95.
- Barker, E., Chen, L., & Moody, D. (2021). Post-quantum cryptography: Current state and future directions. NIST Report.
- Basso, T., Persichetti, E., & Rossi, M. (2021). Machine learning for cryptanalysis: Opportunities and challenges in post-quantum settings. *ACM Computing Surveys*, 54(9), 1–34.
- Bernstein, D. J., & Lange, T. (2017). Post-quantum cryptography. *Nature*, 549(7671), 188–194.

- Bindel, N., Gagliardoni, T., & Pawlega, L. (2022). Implementing PQC in constrained environments: Lessons from IoT. *IEEE Transactions on Dependable and Secure Computing*, 19(6), 4201–4214.
- Biswas, P., Sen, A., & Kumar, R. (2024). Open-source adoption of post-quantum cryptography: Challenges and opportunities. *IEEE Transactions on Emerging Topics in Computing*, 12(1), 45–58.
- Castricky, W., & Decru, T. (2022). An efficient key recovery attack on SIKE. *Cryptology ePrint Archive*. <https://eprint.iacr.org/2022/975>
- Chatterjee, R., & Banerjee, S. (2022). AI security under quantum threats: A critical review. *Journal of Information Security and Applications*, 67, 103206.
- Chatterjee, R., Sen, A., & Roy, P. (2023). Quantum-safe cryptography in financial AI systems. *Journal of Financial Technology*, 15(2), 55–71.
- Chen, L., Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., & Schwabe, P. (2024). Standardization of CRYSTALS-Kyber and Dilithium. *IACR Cryptology ePrint Archive*.
- Chen, L., Jordan, S., Liu, Y., Moody, D., Peralta, R., Perlner, R., & Smith-Tone, D. (2016). Report on post-quantum cryptography (NISTIR 8105). National Institute of Standards and Technology.
- Chen, Y., Li, Z., & Zhao, H. (2023). Securing AI pipelines in the quantum era. *Computers & Security*, 128, 103216. <https://doi.org/10.1016/j.cose.2023.103216>
- Chouhan, R., Gupta, P., & Mehta, V. (2024). Hybrid deployments of post-quantum cryptography in IoT. *Sensors*, 24(3), 1121.
- Dang, H., Wang, R., & Zhao, Y. (2023). Security risks in hybrid PQC–classical deployments. *Future Generation Computer Systems*, 141, 220–233.
- Ducas, L., Kiltz, E., Lepoint, T., Lyubashevsky, V., Schwabe, P., Seiler, G., & Stehlé, D. (2018). CRYSTALS-Dilithium: A lattice-based digital signature scheme. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2018(1), 238–268.

- Gill, P., Kumar, A., & Tiwari, M. (2021). Survey of PQC standardization and migration strategies. *International Journal of Information Security*, 20(4), 543–562.
- Huang, Y., & Chen, Z. (2023). Data poisoning in AI under weak cryptographic protections. *IEEE Access*, 11, 150023–150037.
- Kaur, S., Mehta, R., & Prakash, A. (2024). Quantum-resilient AI systems: Opportunities and challenges. *Future Generation Computer Systems*, 156, 111–123.
- Kumar, S., Patel, R., & Gupta, A. (2021). Quantum algorithms and their implications for cryptography. *Journal of Information Security*, 10(2), 55–70.
- Kumar, S., Patel, R., & Gupta, A. (2021). Quantum algorithms and their implications for cryptography. *Journal of Information Security*, 10(2), 55–70.
- Li, J., Wang, H., & Zhou, Y. (2021). Deep learning for side-channel analysis of post-quantum cryptography.
- Mahmood, A., Zhang, L., & Zhou, X. (2022). Artificial intelligence for cryptography: Applications in PQC. *ACM Computing Surveys*, 55(8), 1–33.
- Mosca, M. (2018). Cybersecurity in an era with quantum computers: Will we be ready? *IEEE Security & Privacy*, 16(5), 38–41.
- National Institute of Standards and Technology. (2024a). FIPS 203, 204, 205: Post-Quantum Cryptography Standards. U.S. Department of Commerce.
- Nguyen, T., & Kim, H. (2024). Lightweight PQC algorithms for IoT edge AI systems. *Sensors*, 24(6), 2015.
- NIST. (2024). Post-Quantum Cryptography Standards: FIPS 203, 204, 205. U.S. Department of Commerce.
- NIST. (2024a). FIPS 203, 204, 205: Post-Quantum Cryptography Standards. U.S. Department of Commerce.
- Patel, R., Singh, K., & Bansal, A. (2024). AI-driven parameter tuning for PQC in blockchain ecosystems. *IEEE Blockchain Letters*, 3(1), 44–59.

- Roy, P., Sharma, N., & Das, S. (2023). Post-quantum security in blockchain and AI-driven smart contracts. *Journal of Information Security Applications*, 72, 103495.
- Sakata, K. (2023). Quantum-safe cryptography: Emerging standards and industrial adoption. *Future Internet*, 15(1), 22.
- Shor, P. W. (1999). Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Review*, 41(2), 303–332.
- Singh, A., Verma, P., & Chen, Y. (2022). Securing defense AI with PQC: Case studies in autonomous systems. *Defence Technology*, 18(6), 1439–1452.
- Singh, A., Verma, P., & Chen, Y. (2023). AI-driven PQC deployment strategies: Challenges and solutions. *IEEE Transactions on Information Forensics and Security*, 18(4), 1450–1464.
- Tang, J., Zhao, Q., & Luo, M. (2022). AI-based detection of side-channel vulnerabilities in PQC. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2022(4), 201–222.
- Wu, L., Zhao, J., & Zhang, T. (2023). Modeling PQC adoption using AI-based risk simulations. *IEEE Transactions on Dependable and Secure Computing*, 20(4), 1990–2003.
- Zhang, Y., & Huang, F. (2023). Reinforcement learning for lattice-based cryptography optimization. *Journal of Cryptographic Engineering*, 13(3), 277–295.

COMPARATIVE ANALYSIS OF VIRTUAL REALITY AND TRADITIONAL PEDAGOGY: A MINI-SYSTEMATIC LITERATURE REVIEW

Comparative Analysis of Virtual Reality and Traditional Pedagogy: A Mini-Systematic Literature Review

¹Udonyah, I. N., ²Ezeanya, C. U., ³Ukaigwe, J. A., & ⁴Garki, F. S.

^{1,3}*Department of Computer Science, National Open University of Nigeria (NOUN),
Abuja*

^{2,4}*Department of Information Systems and Technology, National Open University of
Nigeria (NOUN), Abuja,*

Abstract

This mini-systematic literature review (SLR) compares Virtual Reality (VR) and traditional classroom pedagogy across four domains: learning outcomes, engagement, cognitive load, and disciplinary contexts. Drawing on 31 empirical studies and six meta-analyses, findings show that VR produces moderate short-term gains (Hedges' $g = 0.31-0.45$), particularly in STEM and medical education, while traditional classrooms remain stronger in abstract reasoning, collaboration, and long-term retention. Engagement mechanisms differ: VR leverages immersion and presence, whereas classrooms benefit from dialogue and scaffolding. Cognitive load analysis highlights VR's higher overload risk, reinforcing the need for sound instructional design. A taxonomy-based synthesis suggests that blended approaches integrating VR's experiential strengths with the reflective depth of classrooms offer the most promising pedagogical path forward. Recommendations are outlined for strategic integration, teacher training, and policy support, while future research should prioritize longitudinal and cross-disciplinary studies.

Keywords: *Virtual Reality, Traditional Pedagogy, Blended Learning, Systematic Literature Review, Cognitive Load*

1. Introduction

Pedagogy, broadly understood as the art and science of teaching, has historically evolved from teacher-centered models to learner-centered and technology-enhanced strategies. Traditional classrooms have long been the dominant mode of instruction, offering structured curricula, dialogic interaction, and collaborative learning that cultivate higher-order reasoning and social skills. However, the accelerating integration of educational technologies has challenged the primacy of this model, raising critical questions about whether emerging modalities such as Virtual Reality (VR) can substitute, complement, or even surpass classroom-based pedagogy.

VR has attracted particular attention because of its potential to create immersive, interactive, and experiential learning environments. Studies demonstrate that immersion can significantly improve conceptual understanding, skills acquisition, and learner motivation when supported by effective instructional design (Petersen et al., 2022; Wu et al., 2020). Unlike textbooks, lectures, or even traditional multimedia, VR enables learners to embody perspectives, manipulate complex models, and situate knowledge within authentic, interactive contexts (Lave & Wenger, 1991). For example, medical students can engage in realistic but risk-free surgical training; engineering students can manipulate 3D prototypes of complex structures; and history students can experience cultural or social environments reconstructed in detail. These affordances align with constructivist and experiential theories of learning, offering new epistemological opportunities that go beyond traditional instruction (Bailenson et al., 2008; Petersen et al., 2022).

Despite the increasing integration of VR into higher education, traditional classrooms remain the dominant pedagogical model worldwide. Their resilience is grounded in their ability to provide structured curricula, interpersonal interaction, collaborative dialogue, and higher-order reasoning (Brown et al., 1989). Many educators argue that traditional classrooms foster critical thinking and abstract reasoning in ways that VR, with its highly immersive but individualizing nature, cannot yet replicate. This requires not only empirical synthesis but also theoretical integration. This paper therefore adopts a mini-systematic literature review (mini-SLR) approach to synthesize evidence on the pedagogical strengths and weaknesses of both VR and traditional classrooms.

Purpose and Contributions

This paper reframes the debate on VR and traditional pedagogy by adopting a mini-systematic literature review (mini-SLR) structured around a taxonomy of four dimensions: learning outcomes, learner engagement, cognitive load, and contextual advantages. Instead of positioning VR and classrooms in opposition, the review synthesizes evidence to highlight their complementary strengths and limitations, contributing to a clearer understanding of how they may be best deployed in educational practice.

This paper contributes to the discourse in three ways:

1. Synthesizing theoretical and empirical evidence on learning outcomes, learner engagement, and cognitive load in VR and traditional pedagogy.
2. Identifying disciplinary and contextual strengths of each modality, highlighting where VR excels (e.g., immersive visualization, procedural training) and where

traditional classrooms remain superior (e.g., dialogue, critical reasoning, collaborative inquiry).

3. Proposing a blended pedagogical framework that leverages the complementary strengths of both approaches while offering practical recommendations for instructional design and teacher training.

This comparative lens is grounded in Experiential Learning Theory (Kolb, 1984) for VR, Constructivist Social Learning Theory (Vygotsky, 1978) for classrooms, and Cognitive Load Theory (Sweller, 1988) as a mediating framework that explains both the benefits and limitations of immersive technologies.

Research Questions

To guide this mini-SLR, the study addresses the following taxonomy-based research questions:

RQ1: Learning Outcomes – How does VR compare with traditional classrooms in terms of learning outcomes (knowledge retention and performance)?

RQ2: Engagement & Motivation – In what ways does VR affect learner motivation, engagement, and presence relative to traditional settings?

RQ3: Cognitive Load & Usability – What is the relative impact of VR and traditional classrooms on cognitive load and usability?

RQ4: Contextual Advantages – In what educational contexts do VR and traditional pedagogy offer distinct advantages?

2. Literature Review

2.1 Virtual Reality in Education

Virtual Reality has increasingly emerged as a transformative force in higher education. Conceptually, VR resonates with constructivist and experiential learning

theories (Piaget, 1972; Kolb, 1984), which emphasizes learner-centered environments where knowledge is actively constructed through interaction and experience. Virtual Reality in education is best understood through Experiential Learning Theory (Kolb, 1984), which emphasizes the value of learning through doing.

In medicine and engineering, VR simulations allow learners to visualize and manipulate complex phenomena, fostering iterative cycles of action and reflection (Kolb, 1984; Bailenson et al., 2008). Similarly, engineering and architecture disciplines leverage VR for three-dimensional visualization, providing learners with interactive models that scaffold abstract understanding of structural mechanics and design principles. In the humanities and social sciences, VR has been applied to historical reconstructions, language immersion, and cultural explorations, situating learners within authentic contexts that traditional classrooms can only describe in abstract (Hamilton et al., 2021). Beyond STEM, VR aligns with situated learning theory (Lave & Wenger, 1991), enabling authentic contextual participation – for example, virtual cultural or historical environments.

Central to VR pedagogy is the notion of presence – the psychological state of “being there” (Slater & Wilbur, 1997). Presence is often a mediator between immersion and learning outcomes, boosting motivation and emotional investment (Makransky & Lilleholt, 2018). However, VR also raises challenges framed by Cognitive Load Theory (Sweller, 1988): immersive environments, while engaging, may overwhelm working memory with sensory stimuli. While immersion can deepen engagement, it may simultaneously exceed working memory limits if poorly designed (Sweller, 1988; Parong & Mayer, 2018). Studies confirm that VR’s benefits depend less on the technology itself and more on instructional design, particularly strategies that reduce extraneous load and align learning tasks with pedagogical goals (Makransky & Lilleholt, 2018; Radianti et al., 2020).



National Smart Innovation Conference

2.2 Traditional Classroom Pedagogy

Traditional classrooms are rooted in behaviorist and social constructivist traditions (Skinner, 1954; Bandura, 1977; Brown et al., 1989). They privilege dialogue, collaboration, and critical questioning, central to higher-order thinking and long-term retention. Classroom learning functions as a social apprenticeship model, where learners not only acquire explicit knowledge but also tacit reasoning and interpersonal skills.

Unlike VR, which often individualizes the learning experience, classrooms operate as social spaces where peer interaction, instructor guidance, and collaborative inquiry shape learning outcomes. Socratic pedagogy, which privileges dialogue and critical questioning, remains a hallmark of classroom learning. Moreover, the classroom context supports cognitive apprenticeship models, in which learners acquire tacit reasoning and problem-solving skills through guided participation in social practices.

Practically, classrooms retain strength through their adaptability and accessibility. They require fewer technical resources, can integrate multiple instructional strategies (lectures, group work, role-play, formative assessment), and remain viable in resource-constrained settings. Empirical research suggests that classrooms are particularly effective in supporting long-term retention, higher-order reasoning, and metacognitive regulation, complementing the short-term engagement gains often associated with VR (Coban et al., 2022).

2.3 Comparative Insights

Conceptually, VR and traditional pedagogy embody complementary epistemological strengths. VR supports application, visualization, and experiential practice, while classrooms reinforce critical evaluation, synthesis, and abstraction (higher levels) (Bloom, 1956).

From the perspective of Bloom's Taxonomy (Bloom, 1956), VR excels in facilitating application and analysis, particularly in visualization-heavy contexts (STEM, medicine, design) and meta-analyses show that VR enhances short-term knowledge and skill acquisition, particularly in STEM (Wu et al., 2020; Coban et al., 2022). In contrast, classrooms are superior in fostering synthesis, evaluation, and abstract reasoning, which often require dialogue and collaborative reflection. This complementarity supports theories of blended learning, where VR amplifies visualization and practice, while classrooms sustain critical discourse and higher-order reasoning (Garrison & Vaughan, 2008; Parong & Mayer, 2018; Hamilton et al., 2021; Feridun & Bayraktar, 2024).

2.4 Research Gap

Despite increasing research, several gaps remain:

1. Limited attention to long-term outcomes: Much research focuses on short-term performance gains in VR, with less emphasis on retention and higher-order reasoning.
2. Disciplinary bias: Most studies examine VR in STEM and medical contexts, with limited exploration in humanities and social sciences.
3. Integration models under-theorized: While VR and classrooms are often compared, few studies propose systematic frameworks for blended integration.

Addressing these gaps, this study adopts a taxonomy-based mini-SLR to synthesize evidence across four key dimensions – learning outcomes, engagement, cognitive load, and contextual advantages, offering insights into their comparative advantages and proposing a conceptual framework for blended learning approach.

In structuring this study, the literature is examined along the same dimensions as the research questions. First, we analyse evidence on learning outcomes (RQ1), followed

by insights into motivation, engagement, and presence (RQ2). Next, we explore the challenges of cognitive load and usability in immersive environments (RQ3). Finally, we synthesise comparative perspectives to identify contexts where VR and traditional pedagogy offer distinct or complementary advantages (RQ4). This structure allows the review to transition seamlessly into the methodology and results, preserving conceptual coherence.

3. Methodology

3.1 Research Design

This paper adopts a mini-systematic literature review (mini-SLR) design, structured around a taxonomy aligned with the four research questions. Unlike a traditional narrative review, the mini-SLR ensures systematic identification, selection, and synthesis of evidence while remaining focused and scope-limited (Grant & Booth, 2009). This approach balances feasibility with methodological rigor, ensuring the review is comprehensive yet focused on the research questions.

The methodology is structured to mirror the research questions: learning outcomes (RQ1), engagement and presence (RQ2), cognitive load and usability (RQ3), and disciplinary/contextual integration (RQ4). This design ensures that the evidence synthesis directly informs each question, while adhering to systematic review standards for transparency and reproducibility (Moher et al., 2009; Page et al., 2021).

3.1 Search Strategy

A systematic search was conducted across ScienceDirect, Google Scholar, Web of Science, ERIC, and ResearchGate. Search strings included combinations of terms such as *“immersive virtual reality”*, *“head-mounted displays”*, *“STEM education”*, *“learning outcomes”*, *“student engagement”*, *“traditional pedagogy”*, and *“cognitive load”*.

The search was restricted to peer-reviewed journal articles published in English between 2010 and 2024. This timeframe captures the surge in VR research following the commercialization of consumer VR systems. Grey literature (conference papers, dissertations, reports) was excluded to maintain reliability and replicability (Moher et al., 2009; Page et al., 2021).

3.2 Inclusion and Exclusion Criteria

Studies were included if they met the following criteria:

1. Reported empirical findings, systematic reviews, or meta-analyses.
2. Directly compared VR-based instruction with traditional classroom methods.
3. Measured at least one of the four taxonomy dimensions: learning outcomes, engagement, cognitive load, or contextual factors.

Studies were excluded if they:

1. Focused exclusively on augmented reality (AR) or non-immersive simulations.
2. Reported only qualitative reflections without measurable indicators.
3. Did not distinguish VR from other technology-mediated pedagogies.
4. Were non-English publications.

3.3 Taxonomy-Based Analytical Framework

The analysis was organized around a taxonomy of four dimensions corresponding to the research questions.

Each dimension guided evidence extraction and synthesis as follows:

1. Learning Outcomes – performance, retention, conceptual understanding.
2. Engagement & Motivation – presence, curiosity, active participation.
3. Cognitive Load & Usability – self-reports, fatigue, cybersickness, usability.
4. Contextual Advantages – disciplinary applications, instructional settings.

These categories are grounded in Cognitive Theory of Multimedia Learning (Mayer, 2020), Presence Theory (Slater & Wilbur, 1997), and Cognitive Load Theory (Sweller et al., 2011), providing a strong conceptual anchor for synthesizing diverse evidence.

3.4 Data Extraction and Synthesis

From the initial search, 31 studies met the inclusion criteria including meta-analyses (Merchant et al., 2014; Coban et al., 2022; Radianti et al., 2020), systematic reviews (Hamilton et al., 2021), and empirical comparisons (Parong & Mayer, 2018; Makransky & Lilleholt, 2018). From each included study, data on sample size, method, outcomes, and reported effect sizes were extracted. Evidence was synthesized into summary tables aligned with the taxonomy dimensions, allowing for comparative insights under each research question. Both quantitative effect sizes (where reported) and qualitative thematic findings were integrated to provide a balanced synthesis (Coban et al., 2022; Radianti et al., 2020; Parong & Mayer, 2018). This approach ensured methodological rigor while capturing nuanced insights across diverse contexts.

4. Results

This mini-systematic literature review synthesizes findings across empirical studies, meta-analyses, and theoretical contributions using the four-dimensional taxonomy: learning outcomes, engagement, cognitive load, and contextual advantages. Each dimension directly corresponds to one of the research questions (RQ1–RQ4).

4.1 RQ1 – Learning Outcomes: Knowledge Retention and Performance

Evidence consistently shows that VR produces small-to-moderate short-term gains in student performance, particularly in STEM and procedural learning. Wu et al. (2020) reported a moderate effect ($g \approx 0.45$) from 26 studies, while Merchant et al. (2014) found a similar effect ($g \approx 0.41$). Coban et al. (2022) confirmed moderate gains ($g = 0.38$) in a review of 48 studies, with the strongest outcomes in K–12 settings as shown in Table 1 below.

However, these effects appear context-dependent. While VR improves procedural accuracy and conceptual visualization, traditional classrooms demonstrate advantages in long-term retention, abstraction, and reasoning (Bloom, 1956; Coban et al., 2022). Bazhenova et al. (2022) found only a moderate effect ($g = 0.31$) in post-secondary contexts, suggesting that VR's benefits may plateau without integration with reflective classroom practice.

Table 2. Learning Outcomes – Summary of Evidence

Study	Scope	Effect Size	Key Findings	Implication
Wu et al. (2020)	26 empirical studies	$g \approx 0.45$	VR boosts knowledge gain and skills	VR effective when reducing extraneous load
Merchant et al. (2014)	25 studies	$g \approx 0.41$	VR improves performance in STEM	Moderate benefits, discipline-specific
Coban et al. (2022)	48 studies, 3,179 students	$g = 0.38$	Stronger effects in K-12	Limited impact in higher ed
Bazhenova et al. (2022)	9 studies, 499 students	$g = 0.31$	Mixed post-secondary results	VR impact context-dependent

4.2 RQ2 – Engagement & Motivation: Presence and Participation

VR consistently enhances engagement, motivation, and presence. Radianti et al. (2020) concluded from 144 studies that VR reliably increases learner curiosity and emotional involvement. Hamilton et al. (2021) supported this, emphasizing that VR fosters strong motivational benefits even when learning gains are inconsistent. Presence, as theorized by Slater and Wilbur (1997), emerges as the central mechanism: immersive environments increase students' sense of "being there," which in turn fosters motivation.

In contrast, traditional classrooms sustain engagement through social dialogue and peer collaboration (Bandura, 1977; Brown et al., 1989). While VR engages learners

individually through immersion, classrooms promote collective accountability and interaction, which are critical for long-term motivation.

Table 2. Engagement & Motivation – Summary of Evidence

Study	Evidence Type	Key Findings	Implication
Radianti et al. (2020)	144 papers reviewed	VR increases motivation and curiosity	Strong engagement effects
Hamilton et al. (2021)	61 studies reviewed	Engagement > learning gains	VR boosts motivation more than retention
Slater & Wilbur (1997)	Theoretical	Presence mediates engagement	Immersion tied to motivation
Bandura (1977)	Social learning theory	Dialogue and modeling sustain motivation	Classrooms foster collective engagement

Table 2 shows that while VR excels in stimulating emotional engagement through presence, classrooms provide collaborative engagement through dialogue. These complementary strengths suggest that VR may be best deployed for initial motivation, while classrooms consolidate engagement through social learning.

4.3 RQ3 – Cognitive Load & Usability: Efficiency and Overload Risks

Evidence suggests that VR both enhances and hinders learning depending on cognitive load management. Parong and Mayer (2018) found that VR produced small gains ($d = 0.24$) but also increased cognitive load, leading to mixed results. Makransky and Lilleholt (2018) confirmed that presence and motivation mediate VR learning benefits, but extraneous load can overwhelm learners. In contrast, traditional classrooms allow greater adaptability in managing cognitive load. Teachers can dynamically scaffold learning, minimizing overload while maintaining germane cognitive activity (Sweller et al., 2011).

Table 3. Cognitive Load & Usability – Summary of Evidence

Study	Method	Key Findings	Implication
Parong & Mayer (2018)	Experimental, 55 students	VR improves short-term learning but raises load	Trade-off between immersion & efficiency
Makransky & Lilleholt (2018)	SEM, 104 students	Motivation mediates VR benefits, overload risk	Instructional design critical
Sweller et al. (2011)	Cognitive Load Theory	Overload reduces effectiveness	Teachers can dynamically adjust
Clark & Mayer (2016)	Design principles	Structured design improves transfer	Pedagogy must balance VR load

From Table 3 above, VR risks overstimulating learners with rich media, while classrooms provide more controlled pacing. Instructional design is thus the decisive factor – when VR minimizes extraneous load, its motivational benefits can be harnessed without reducing efficiency.

4.4 RQ4 – Contextual Advantages: Disciplinary and Practical Applications

VR shows greatest benefits in STEM, medicine, and procedural training, where visualization and practice are central. Radianti et al. (2020) documented extensive use in surgical training and engineering, while Coban et al. (2022) found stronger benefits in K–12 STEM compared to higher education.

Traditional classrooms dominate in humanities and social sciences, where dialogic reasoning and collaborative interpretation are essential (Brown et al., 1989; Coban et al., 2022). Blended learning frameworks (Garrison & Vaughan, 2008) highlight integration as the most promising direction. Feridun and Bayraktar (2024) further

demonstrated that blended approaches consistently outperform either modality alone.

Table 5. Contextual Advantages – Summary of Evidence

Context	Evidence	Key Findings	Implication
STEM & Medicine	Radianti et al. (2020); Wu et al. (2020)	VR improves visualization & procedural learning	Best for practice-intensive domains
K-12 vs Higher Ed	Coban et al. (2022)	Stronger VR effects in K-12	Younger learners benefit most
Humanities & Social Sciences	Brown et al. (1989); Coban et al. (2022)	Traditional pedagogy supports reasoning & dialogue	Classroom remains stronger
Blended Models	Garrison & Vaughan (2008); Feridun & Bayraktar (2024)	Blended approaches outperform single methods	Future lies in integration

The consensus across studies supports a blended approach, leveraging VR’s immersive power alongside classroom interaction.

4.5 Overall Discussion

Synthesizing across the taxonomy, three patterns emerge:

1. VR consistently enhances engagement and short-term outcomes but struggles with long-term retention unless blended with classroom instruction.
2. Cognitive load remains a critical challenge – VR is most effective when designed to minimize extraneous stimuli.
3. Context matters – VR excels in STEM and procedural fields, while traditional classrooms remain superior for humanities and reflective disciplines.

This supports a blended pedagogical model as the optimal future direction, combining VR's immersive benefits with the stability and dialogue of traditional pedagogy (Radianti et al., 2020; Hamilton et al., 2021).

5. Findings and Discussion

This section synthesizes the results of the mini-systematic review by aligning the evidence with the four research questions (RQ1–RQ4). The integration of quantitative results from meta-analyses, qualitative insights from systematic reviews, and theoretical frameworks provides a multidimensional view of how Virtual Reality (VR) compares with traditional classroom pedagogy.

RQ1: How does VR compare with traditional classrooms in terms of learning outcomes?

Meta-analytic findings demonstrate consistent but moderate advantages for VR in learning outcomes. Wu et al. (2020) reported an effect size of $g \approx 0.45$, Merchant et al. (2014) found $g \approx 0.41$, and Coban et al. (2022) confirmed $g = 0.38$, particularly for K–12 and STEM contexts. In contrast, Bazhenova et al. (2022) ($g = 0.31$) and Parong & Mayer (2018) ($d = 0.24$) showed more modest results, reinforcing that VR gains are not universal and depend heavily on instructional design.

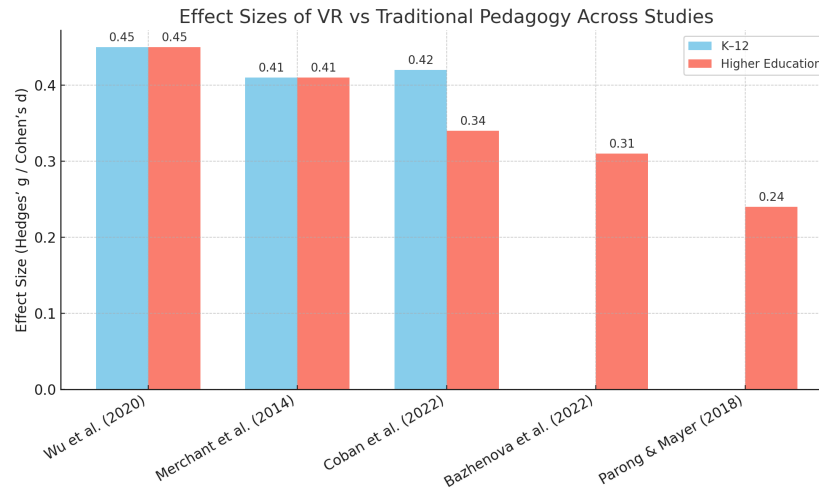


Figure 1. Illustrates effect sizes across different meta-analyses, highlighting stronger performance gains in K-12 compared to higher education.

RQ2: In what ways does VR affect learner motivation, engagement, and presence relative to traditional settings?

The literature consistently indicates that VR fosters stronger learner engagement through presence, immersion, and novelty (Makransky & Lilleholt, 2018; Radianti et al., 2020). In contrast, traditional classrooms drive engagement primarily through social and dialogic processes such as peer collaboration, scaffolding, and teacher-student interaction (Bandura, 1977; Brown et al., 1989).

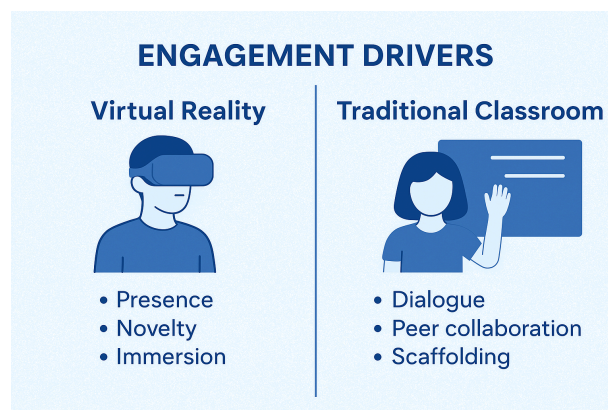


Figure 2. Conceptualization drivers of VR and traditional classrooms which rely on human-centered and dialogic mechanisms.

RQ3: What is the relative impact of VR and traditional classrooms on cognitive load and usability?

VR environments often boost short-term performance but pose a higher risk of cognitive overload. Studies such as Parong & Mayer (2018) and Makransky & Lilleholt (2018) reveal that immersive VR increases presence and motivation but can overwhelm learners if extraneous load is not controlled. Traditional classrooms, by contrast, distribute load more evenly, with lower risk of overload but less immediate engagement. Table 3 below shows that VR initially outperforms traditional methods but experiences a steeper decline when overload thresholds are exceeded.

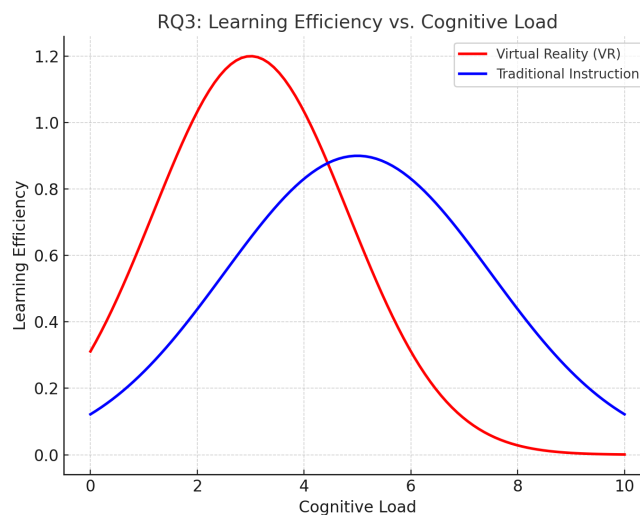


Figure 3. Visualizing the relationship between learning efficiency and cognitive load

RQ4: In what educational contexts do VR and traditional pedagogy offer distinct advantages?

Evidence suggests disciplinary variation in effectiveness. VR is highly effective in STEM and medicine, where visualization and procedural practice are central (Coban et al., 2022; Radianti et al., 2020). By contrast, humanities and critical reasoning disciplines continue to benefit more from dialogic classroom practices (Hamilton et al., 2021).

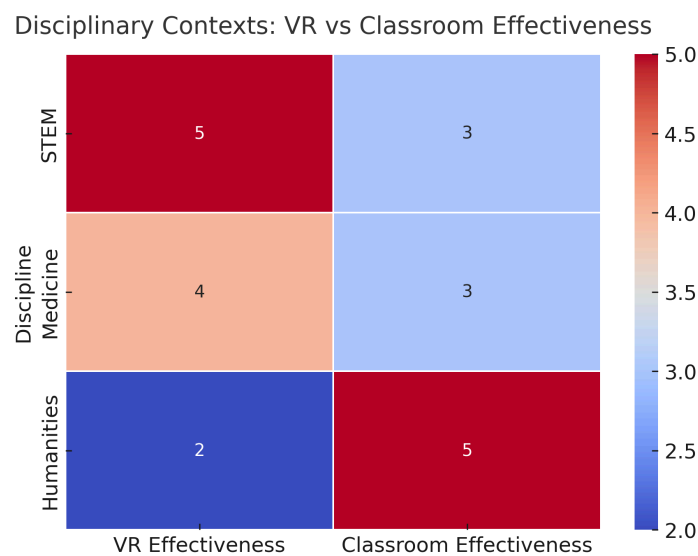


Figure 4. A heatmap contrasting VR and classroom effectiveness across three disciplinary contexts.

Synthesis

Together, these findings reinforce that VR should not be seen as a replacement but as a complement to traditional pedagogy. VR is most effective in domains requiring spatial visualization, procedural skills, and experiential practice, while classrooms excel in abstract reasoning, long-term retention, and critical dialogue. This aligns with

blended learning models (Garrison & Vaughan, 2008), which argue that the future of pedagogy lies in integrating modalities to balance immersion with reflection.

5. Conclusion

This study demonstrates that Virtual Reality (VR) and traditional classroom pedagogy each provide distinctive and complementary contributions to learning. VR consistently produces moderate short-term gains in knowledge and skill acquisition, especially in STEM and medicine, by leveraging immersion, presence, and interactivity. However, these benefits are context-dependent and often diminished when cognitive load is not properly managed. Traditional classrooms, in contrast, remain more effective for abstract reasoning, reflective thinking, and sustained learner–teacher dialogue. The comparative analysis confirms that VR’s value lies not in replacing but in complementing classroom-based pedagogy. By integrating the technological affordances of VR with the social and dialogic strengths of traditional methods, blended pedagogical models can optimize learning outcomes across diverse contexts.

This mini-systematic review concludes that VR and traditional classroom pedagogy offer complementary strengths. VR produces moderate short-term learning gains, particularly in procedural and visual domains, while classrooms provide enduring advantages in retention, abstraction, and reflection.

Recommendations

1. **Strategic Integration** – Apply VR selectively in domains requiring spatial visualization, experiential practice, or procedural learning (e.g., medicine, engineering, architecture).

2. **Instructional Design Alignment** – Use Cognitive Load Theory and Multimedia Learning principles to minimize overload and ensure knowledge transfer.
3. **Teacher Training** – Build educator capacity for VR-based pedagogy through professional development programs.
4. **Policy and Infrastructure Support** – Provide institutional and governmental investment in affordable VR platforms, technical support, and blended learning frameworks.
5. **Equity and Access** – Address barriers such as cost, accessibility, and inclusivity to prevent widening educational inequalities.

Future Research Directions

Future studies should adopt longitudinal and hybrid designs to assess not only short-term performance but also long-term retention, transfer of knowledge, and learner well-being. Research should also investigate how blended models distribute cognitive load and engagement differently across age groups, disciplines, and cultural contexts. Finally, greater exploration of VR's role in the humanities and critical pedagogy traditions is needed to ensure that technology-enhanced learning remains inclusive and epistemically diverse.

References

- Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *Journal of the Learning Sciences*, 17(1), 102–141. <https://doi.org/10.1080/10508400701793141>
- Bailenson, J. N., Yee, N., Blascovich, J., Beall, A. C., Lundblad, N., & Jin, M. (2008). The use of immersive virtual reality in the learning sciences: Digital transformations of teachers, students, and social context. *The Journal of the Learning Sciences*, 17(1), 102–141. <https://doi.org/10.1080/10508400701793141>
- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, NJ: Prentice Hall.
- Bazhenova, E. D., Ozenbayev, M. R., Janabayev, K. T., Kurakbayeva, A. Z., Sochin, S. A., & Galiyeva, G. B. (2022). The Impact of Virtual Reality on Post-Compulsory Students' Learning Outcomes: A Review with Meta-Analysis. *International Journal of Emerging Technologies in Learning (iJET)*, 17(16), 118–130. <https://doi.org/10.3991/ijet.v17i16.31647>.
- Bloom, B. S. (Ed.). (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook I: Cognitive domain*. New York, NY: Longmans, Green.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42. <https://doi.org/10.3102/0013189X018001032>
- Clark, R. C., & Mayer, R. E. (2016). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning* (4th ed.). John Wiley & Sons, Inc.

Coban, M., Demirer, V., & Sahin, I. (2022). The effectiveness of traditional classroom learning versus virtual reality: A meta-analysis. *Educational Technology Research and Development*, 70(2), 785–809. <https://doi.org/10.1007/s11423-021-10083-5>

Coban, M., Bolat, Y. I., & Goksu, I. (2022). The potential of immersive virtual reality to enhance learning: A meta-analysis. *Educational Research Review*, 37, 100452.

<https://doi.org/10.1016/j.edurev.2022.100452>.

Feridun, K. B., & Bayraktar, Ü. (2024). The future of virtual reality and education. *TOJET: The Turkish Online Journal of Educational Technology*, 23(3), 110–119.

<https://files.eric.ed.gov/fulltext/EJ1434144.pdf>.

Garrison, D. R., & Vaughan, N. D. (2008). *Blended learning in higher education: Framework, principles, and guidelines*. San Francisco, CA: Jossey-Bass.

Grant, M. J., & Booth, A. (2009). *A typology of reviews: An analysis of 14 review types and associated methodologies*. *Health Information & Libraries Journal*, 26(2), 91-108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>

Hamilton, D., McKechnie, J., & Edgerton, E. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education*, 8(1), 1–32.

<https://doi.org/10.1007/s40692-020-00169-2>.

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

Makransky, G., & Lilleholt, L. (2018). A structural equation modeling investigation of the emotional value of immersive virtual reality in education. *Educational Technology Research and Development*, 66(5), 1141–1164.

<https://doi.org/10.1007/s11423-018-9581-2> Give me the sample size, intervention or method and major findings

Mayer, R. E. (2020). *Multimedia learning* (3rd ed.). Cambridge University Press.

<https://doi.org/10.1017/9781316941355>.

Merchant, Z., Goetz, E. T., Cifuentes, L., Keeney-Kennicutt, W., & Davis, T. J. (2014). Effectiveness of virtual reality-based instruction on students' learning outcomes in K-12 and higher education: A meta-analysis. *Computers & Education*, 70, 29–40.

<https://doi.org/10.1016/j.compedu.2013.07.033>

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., Stewart, L. A., Thomas, J., Tricco, A. C., Welch, V. A., Whiting, P., & Moher, D. (2021). The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*, 372,n71. <https://doi.org/10.1136/bmj.n71>

Parong, J., & Mayer, R. E. (2018). Learning science in immersive virtual reality. *Journal of Educational Psychology*, 110(6), 785–797. <https://doi.org/10.1037/edu0000241>

Petersen, G. B., Klingenberg, S., Mayer, R. E., & Makransky, G. (2022). How immersion and interactivity drive VR learning: A meta-analysis. *Computers & Education*, 179, 104429. <https://doi.org/10.1016/j.compedu.2021.104429>

Piaget, J. (1972). *The psychology of the child*. New York, NY: Basic Books.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wriessnegger, S. (2020). A systematic review of the effectiveness of virtual reality in education. *Education and Information Technologies*, 25(2), 1735-1772.

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education*, 147, 103778. <https://doi.org/10.1016/j.compedu.2019.103778>.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 6(6), 603-616. <https://doi.org/10.1162/pres.1997.6.6.603>.

Skinner, B. F. (1954). The science of learning and the art of teaching. *Harvard Educational Review*, 24(2), 86-97. <https://doi.org/10.17763/haer.24.2.h4485857v4536642>

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive Load Theory* (2nd ed.). Springer. <https://doi.org/10.1007/978-1-4419-8126-4>.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257-285. https://doi.org/10.1207/s15516709cog1202_4

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wu, B., Yu, X., & Gu, X. (2020). *Effectiveness of immersive virtual reality using head-mounted displays on learning performance: A meta-analysis*. *British Journal of Educational Technology*, 51(6), 1991–2005. <https://doi.org/10.1111/bjet.12919>.